INCREASING THE VALUE OF SOYBEAN THROUGH BREEDING FOR IMPROVED SEED PROTEIN CONTENT AND OPTIMIZING PRODUCTION PRACTICES

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Landon Linn Ries

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. James Orf, Dr. Seth Naeve, Advisors

April, 2018

Acknowledgements

First and foremost, I thank my wife Tara for supporting me throughout my academic and professional career. Without your support and stability, most assuredly, this would not be possible. In a close second place I thank my son Lucas; for giving me perspective to the priorities of life. I thank my parents and sisters for supporting me through my life in general. Without their guidance, I would not be the person I am today.

I thank Drs. Seth Naeve and James Orf for giving me the opportunity to work with their programs and for directing me through my research. I thank Dr. Robert Stupar and his lab for providing me with collaborative opportunities to enable me to learn genomic strategies with applied crop improvement a central focus. Collectively, the knowledge and skills I have gained while working with Drs. Orf, Naeve, and Stupar and their respective research programs provides me with a solid foundation for the rest of my career as a crop scientist. I thank the United Soybean Board and the Minnesota Soybean Research and Promotion Council for their funding of this research.

I thank Dr. Jill Miller-Garvin, Dimitri Von Ruckert, Raphael Echenique, Phil Shaus, Art Killam, Darcy Weston, Gerald Decker, Young Tsi Bolon, Jean-Michel Michno and others within the Soybean Breeding, Soybean Production/Physiology, and Soy Genomics Labs for their advice, knowledge, and assistance throughout my research. Special thanks to my graduate committee: Drs. Robert Stupar, James Anderson, and Nevin Young for their input in my graduate curriculum and research feedback.

Dedication

I dedicate this document to my Dad, Linn, and Mom, Sharla. The environment for which they provided me while growing up and their never-ending love and support made all things possible.

Table of Contents

List of Tables

Chapter 2

Chapter 3

List of Figures

Chapter 2

Chapter 3

Chapter 1 - Introduction

The value of the United States soybean (*Glycine max* (L. Merr) crop is greater than $40B USD (2016; SoyStats, USDA-NASS). Much of this value is due to the protein content within the seed which is used as a primary component in animal feeds as well as human foodstuffs (Wilson, 2004; Medic et al., 2014). The value of soybean is inherently linked to the composition of the seed and therefore soybean seed with lower protein content has lower processed value (Brumm and Hurburgh, 1990; 2006).

Soybean breeders and agronomists have two fundamental methods for improving the value of the soybean crop to producers and processors:

1. Increase yield per unit area and/or
2. Increase the intrinsic value of the crop

Yield improvement can be accomplished through genetic improvement by typical plant breeding methodologies which, historically, have proven successful, particularly in the last four decades (Fox et al., 2013; Rincker et al., 2015). Incorporating pest resistance genes is another strategy soybean breeders may implement to protect the intrinsic yield potential of cultivars (Concibido et al., 2004; Palmer et al., 2004). Moreover, yield improvement may be made through optimization of agronomic and/or cultural practices. Earlier investigators have evaluated optimum planting dates (Anderson and Vasilas, 1985; Beuerlein, 1988; DeBruin and Pedersen, 2008a; Egli and Cornelius, 2009; Elmore, 1990; Grau et al., 1994; Lee et al., 2008; Lueschen et al., 1992; Oplinger and Philbrook, 1992; Pedersen and Lauer, 2004; Wilcox and Frankenberger, 1987), row spacing (Ablett et al., 1991; Alessi and Power, 1982; Beuerlein, 1988; Bullock et al., 1998; Cooper,

1977; DeBruin and Pedersen, 2008b; Devlin et al., 1995; Egli, 1988; Egli, 1994; Elmore, 1991; Ethredge et al., 1989; Grau et al., 1994; Holshouser and Whittaker, 2002; Janovicek et al., 2006; Kratochvil et al., 2004; Lee, 2006; Lueschen et al., 1992; Oplinger and Philbrook, 1992; Pedersen and Lauer, 2003; Taylor, 1980; Weber et al., 1966; Wiggans, 1939)), and plant population density (Ablett et al., 1991; Beuerlein, 1988; Carpenter and Board, 1997; Costa et al., 1980; Cox and Cherney, 2011; DeBruin and Pedersen, 2008a,b; Devlin et al., 1995; Edwards and Purcell, 2005; Egli, 1988; Elmore, 1991; Elmore, 1998; Ethridge et al., 1989; Lee et al., 2008; Leffel and Barber, 1961; Lehman and Lambert, 1960; Lueschen and Hicks, 1977; Parks et al., 1982; Kratochvil et al., 2004; Oplinger and Philbrook, 1992; Pedersen and Lauer, 2002; Walker et al., 2010; Weber et al., 1966; Wells, 1991). Optimum management of soil fertility and reducing pests has been additional topics investigated for soybean yield production.

The value of the soybean crop can be increased most notability though the alteration of seed composition traits (Updaw and Nichols, 1980; Orf and Helms, 1994; Brumm and Hurburgh, 2006); increasing the concentration of higher value fractions via contemporary plant breeding processes or through specific gene modification (e.g. mutagenesis) (Wilson, 2004; Fehr, 2007; Bolon et al., 2014). Total seed protein content is a primary compositional component for which the value of soybean seed is derived (Brumm and Hurburgh, 1990). It has been well documented that the Northern and Western U.S. produces soybean seeds that are lower in protein content relative other regions across the U.S. (Hurburgh 1994; Rotundo et al., 2016; Breene et al., 1988; Yaklich et al., 2002). Hurburgh (1994) showed a protein content decrease of

approximately one percent for soybeans grown in Minnesota, South Dakota, and North Dakota across an eight year study. Rotundo et al. (2016) further demonstrated U.S. protein trend by spatially modeling geo-referenced soybean samples from producer's fields across the U.S. Over the eight years included in the study, a strong protein decreasing trend from north to south is evident. In essence the situation in Minnesota and other Northern portions of the U.S. is the intrinsic environment is not conducive for "high" protein soybean production with current commodity soybean genetics and through typical cross breeding and selection strategies for yield. Taken together, it is clear that a concerted effort to increase protein content in these regions will be necessary to improve the market competitiveness for producers and processors in this northern geography. Given the importance of seed protein content to the value of the crop, there has been much research investigating the phenotypic variation and genetic architecture that exists within the improved and plant accession germplasm pools. While much variation exists, particularly in the USDA soybean germplasm collection, developing higher protein lines while conserving or improving yield has been challenging due to the well documented negative correlation between seed protein content and yield (Johnson et al., 1955; Hanson et al., 1961; Shannon et al., 1972; Brim and Burton, 1979; Sebern and Lambert 1984; Wilcox and Cavins, 1995; Hartwig and Hinson, 1972; Wehrmann et al., 1987, Thorne and Fehr, 1970; Cober and Voldeng, 2000; Helms and Orf, 1998; Simpson and Wilcox, 1983; Chung et al., 2003; Sebolt et al., 2000).

This dissertation research aims to merge the two fundamental themes described above. Chapter 2 focuses on breeding for improved seed protein content while maintaining yield, Chapter 3, focuses on optimizing soybean yield based on seed/plant density. For Chapter 2, A project was initiated by the University of Minnesota Soybean Breeding Program to introgress a native soybean allele from a plant introduction into an elite soybean cultivar adapted Minnesota environments to ultimately increase seed protein content for northern U.S. environments. Through several cycles of phenotypic selection and backcrossing, an advanced backcross population that is genetically similar to the adapted cultivar yet varies for seed protein content was created. This population was phenotypically characterized for agronomic and seed composition traits across Minnesota environments and the underlying genetics contributing to the phenotypic variation was explored. Lines from the population were found to have similar yield to the adapted, recurrent parent with significantly increased seed protein content. The primary genetic locus responsible for elevating protein content was located at a position previously identified for contributing to increased soybean seed protein content. While the region is not novel for increasing protein content, the lines created through the advanced backcross process demonstrate that yield maintenance with elevated seed protein is possible through introgressing the locus into adapted germplasm. Chapter 3 concentrates on increasing/optimizing soybean grain yield per unit area. It was explored through investigating soybean yield response to varying plant densities across the United States from 30.2° to 47.8° N latitude. Soybean seeding rate decisions are made annually by soybean producers, and while the topic has been heavily investigated since the

introduction of soybean to the United States, it has been studied on a local level often specific to soil types, the soybean cultivars used, and field plot techniques implemented. While these results are relevant to the specific environments tested, evaluation of trends using a consistent experimental design across a wide range of latitudes had not been carried out.

Chapter 2 – Developing Soybean Lines with Increased Seed Protein Content and High Yield for Northern U.S. Environments through the Introgression of High Protein from PI 153296

Outline

The protein within the soybean seed provides much of the value of the commodity. Soybean produced in the North Central and Western portions of the United States experience lower seed protein compared to soybean grown in other parts of the country. Given the value of the grain is driven, in part, by protein content, producers and processors in this region are adversely affected. Increasing soybean seed protein content through breeding has shown to be effective, however, often at the cost of overall grain yield and oil content. To address this issue, the University of Minnesota soybean breeding program initiated a project to increase seed protein content through introgressing high seed protein alleles from the *Glycine max* plant introduction PI153296 into the Minnesota adapted cultivar 'Evans'. Through four generations of successive backcrossing, protein content was increased and yield was maintained. Although oil content was diminished, this combination of high protein and high yield within breeding lines is appealing within the target geography. QTL mapping was carried out to identify the genomic regions contributing to seed composition traits and yield. The primary driver of seed protein was found to derive from the previously reported QTL region on chromosome 20. Enabled through the genetic characterization of the USDA-GRIN soybean collection with the Soy50KSNP assay and previous genomic characterizations of this region, a comparison between recurrent parent and donor line indicate that the polymorphic segments in the region are contained to two of the five delineated haplotype blocks.

Introduction

The value of the United States soybean (*Glycine max* (L. Merr) crop is greater than $40B USD (2016; SoyStats, USDA-NASS). The states of Minnesota, South Dakota, and North Dakota account for nearly 20% of this $40B. Much of this value is due to the protein content within the seed which is used as a primary component in animal feeds as well as human foodstuffs (Wilson, 2004; Medic et al., 2014). The value of soybean is inherently linked to the composition of the seed and therefore soybean seed with lower protein content has lower processed value (Brumm and Hurburgh, 1990; 2006). It has been well documented that the Northern and Western U.S. produces soybean seeds that are lower in protein content relative other regions across the U.S. (Hurburgh 1994; Rotundo et al., 2016; Breene et al., 1988; Yaklich et al., 2002). Hurburgh (1994) showed a protein content decrease of approximately one percent for soybeans grown in Minnesota, South Dakota, and North Dakota across an eight year study. Rotundo et al. (2016) further demonstrated U.S. protein trend by spatially modeling geo-referenced soybean samples from producer's fields across the U.S. Over the eight years included in the study, a strong protein decreasing trend from south to north is evident.

While producer's value soybean in terms of mass per unit area produced, the value of the soybean grain is ultimately established by the end-user (Updaw and Nichols, 1980; Orf and Helms, 1994; Brumm and Hurburgh, 2006). Soybean processors are generally focused on crude protein and oil; however, the value of a unit of soybean grain can be discounted when 48% protein meal cannot be generated; this occurs when the soybean grain feedstock delivered to the processing facility falls below 35% protein on a

13% moisture basis (Brumm and Hurburgh, 2006; Wilson, 2004). In most geographies across the United States, a near exclusive focus on yield will result in the greatest processed value as well as the greatest revenue to the producer because, protein content in particular, does not drop below the 48% protein threshold (Brumm and Hurburgh, 2006). As mentioned previously, soybeans produced in the North and West portions of the United States (e.g. Minnesota, North Dakota, and South Dakota) do not meet this threshold at times; thus the price processors are willing to pay in these areas is less than that of other areas creating a market scenario where producers in this region are paid differentially for the crop compared to producers in other soybean growing regions. Thus, increasing protein content in this region while maintaining yield will increase processor's and producer's profitability.

The general, macro trend in decreasing soybean seed protein content from south to north across the U.S. can be attributed to the intrinsic differences in the growing environments across this area as well as specific cultivar genetic factors (Medic et al., 2014). Micro environment and production factors play a role and contribute toward variability within a defined region. These factors include planting date (Rowntree et al., 2013; Helms et al., 1990), specific cultivar selection by the producer, soil water status (Dornbos et al., 1989; Dornbos and Mullen, 1992; Rotundo and Westgate, 2010; Specht et al., 2001), temperature (Dornbos and Mullen, 1992; Piper and Boote, 1999; Pipolo et al., 2004), and soil fertility (Wilson et al., 2014; Ham et al., 1975; Nakasathien et al., 2000; Ray et al., 2006), and spatial topography (Kravchenko and Bullock, 2002).

In addition to the production and environmental effects on soybean seed composition, the historic trend in soybean cultivars released since the 1930's indicates that protein content in maturity group II and III cultivars is decreasing by approximately 0.2 kg ha^{-1} per year (Rowntree et al., 2013; Wilson et al., 2013). A similar trend, protein content has been decreasing in cultivars released by maturity group II and maturity group III breeders, although not as clearly defined as Rowntree et al. (2013) and Wilson et al. (2013), was identified by Mahmoud et al. (2006). A study by Voldeng et al. (1997) demonstrated that protein content in maturity group 0 and earlier cultivars have been following a similar trend. Taken together, the Northern and Western United States region has an environment that results in decreased protein content relative to other regions and newly released cultivars have lower protein content than older cultivars. Environmental characteristics cannot be changed; production practices may offer an opportunity to increase protein content, although, some practices come at a decrease in yield which would be unfavorable to producers (e.g. delayed planting dates resulting in decreased yield); thus, the greatest opportunity to increase protein content in this region is through the development of higher protein content cultivars that maintain yield.

While much variation exists within the soybean germplasm collection (USDA-GRIN) for seed composition traits; developing high protein, high yielding lines has proven to be challenging due to the well documented negative correlation between seed protein content and seed yield (Johnson et al., 1955; Hanson et al., 1961; Shannon et al., 1972; Brim and Burton, 1979; Sebern and Lambert 1984; Wilcox and Cavins, 1995; Hartwig and Hinson, 1972; Wehrmann et al., 1987, Thorne and Fehr, 1970; Cober and

Voldeng, 2000; Helms and Orf, 1998; Simpson and Wilcox, 1983; Chung et al., 2003; Sebolt et al., 2000). Breeding schemes designed to combine high yield and high protein have been met with varied success depending on the specific breeding strategy implemented and the germplasm investigated.

Bi-parental crosses with a high-yield, adapted parent crossed with a high-protein parent have been shown to capture protein contents greater than the elite parent, although yield performance equaling that of the elite parent has been rare (Simpson and Wilcox, 1983; Seabern and Lambert, 1984; Shannon et al., 1972; Helms and Orf, 1997; Chung et al., 2003). Thorne and Fehr (1970) demonstrated 3-way crosses were more favorable than bi-parental crosses when aiming to increase protein content. Several investigators have used backcrossing as a means to capture high protein alleles from high protein parents while maintaining the presence of high-yielding, adapted alleles from the elite recurrent parent (Hartwig and Hinson, 1972; Wehrmann et al., 1987; Wicox and Cavins, 1995; Cober and Voldeng, 2000). Wilcox and Cavins (1995) used two cycles of successive backcrossing to create and recover a line with high-protein that did not differ in yield compared to the recurrent parent. Similarly, Cober and Voldeng (2000) demonstrated that high-protein alleles from an improved high-protein parent can be transferred while recovering high-yield alleles in either single cross or backcross populations. Thus, while the majority of the literature has shown a negative seed yield: seed protein content correlation; it is evident that the magnitude of the negative correlation does not prevent the development of high-yielding, higher-protein cultivars.

In addition to the breeding effort that has taken place for improving soybean protein content, there is been a corresponding effort to understand the genetic control of seed protein. Through the identification of genes/QTL contributing to protein content, breeders may use marker assisted selection methodologies to improved seed protein content in their germplasm. To date, there have been 150+ reported QTL for soybean seed protein content (SoyBase, the USDA, ARS Soybean Genetics and Genomics Database). Most notably, the QTL reported on LG I (chromosome 20; Gm20) has proven to be the largest effect and consistent over environments and germplasm evaluated (Nichols et al., 2006; Sebolt et al., 2000; Diers et al., 1992; and Chung et al., 2003; Hwang et al., 2014; Vaughn et al., 2014; Bandillo et al., 2015). This QTL has been designated sqSeed protein-003 by the Soybean Genetics Committee (SoyBase) and appears to originate from both *G. soja* (e.g. PI468916, Diers et al., 1992) and *G. max* (e.g. PI437088A, Chung et al., 2003) accessions in the germplasm collection; Thus it is possible that the chromosome 20 QTL is in many of the high protein lines contained in the USDA-GRIN germplasm collection. Hwang et al. (2014) demonstrated the power of association mapping in the context of selective genotyping high-protein and average-protein accessions from the germplasm collection. Using 298 accessions from the soybean germplasm collection, which exhibited a wide range in seed protein and oil concentrations, they found many of the previously reported QTL identified by the classical linkage analyses referenced previously, including strong associations on chromosome 20. They were able to localize the causative gene(s) region down from the 8.4 Mbp region reported by Bolon et al., (2010) to a 2.4 Mbp region within the 8.4 Mbp

segment. Within this region Hwang et al. (2014) identified four markers position on chomosome 20 with the SoySNP50K array (Song et al., 2013) that were in strong linkage disequilibrium (LD) with the likely causative gene(s). Vaughn et al. (2014) further refined the region using association mapping with a set of 934 MG V accessions; in their analyses they identified a 1 Mbp segment between approximately 31-32 Mbp. This region is slightly downstream of the 2.4 Mbp segment identified by Hwang et al. (2014). Bandillo et al. (2016), providing the most comprehensive association analysis for soybean seed protein and oil content to date, evaluated the entire germplasm collection utilizing all non-redundant accessions with genotype and phenotypic data in the database. The authors (Bandillo et al., 2016) delineated five haplotype blocks spanning the 8.4 Mbp region defined by Bolon et al. (2010), the majority of the significant maker associations were located in the region defined by Vaughn with the most significant marker located at 31,243,150bp.

Prior to much of the knowledge on the understanding of the genetic control of seed protein content was discovered, the University of Minnesota Soybean Breeding Project initiated a long term successive phenotype based backcrossing strategy to incorporate the high protein from the *G. max* accession PI 153296 (~53 % protein, dry weight basis; Bernard et al., 1998) to cv. Evans (~41% protein, dry weight basis; Lambert and Kennedy, 1975; PI 548560 NPGS-GRIN ID). The original objective at the onset of the project 20+ years ago was to develop germplasm that captures the yield performance of ‘Evans’ while elevating protein content. In light of the current situation with diminishing protein contents in newly developed soybean cultivars is compounded with

the intrinsic environmental factors that lead to the North and Western U.S. soybean growing region having lower seed protein content relative to other US geographies (Rowntree et al., 2013; Rotundo, et al., 2016); we believe this mission continues to be of great interest to the soybean improvement community. The results of this project may shed some additional light on the efficacy of the chromosome 20 protein QTL in Northern US environments, if it is ultimately responsible for driving protein gains in this material. The primary the objectives of this study were to: 1. Characterize the BC_4F_5 population created from the recurrent parent Evans crossed with the high protein donor PI153296 across eight Minnesota site-years; a focus on yield recovery and protein gain and stability will be assessed, and 2. To investigate the genetic control for the protein content and yield within the population and compare results to previous reports.

Materials and Methods

Population Development

Maturity Group 0 cultivar 'Evans' (Lambert and Kennedy, 1975), adapted to Central Minnesota environments, averaging 41% protein content on a dry basis, was crossed to the Maturity Group 00 high-protein *G. max* accession PI 153296 (Bernard et al., 1998), averaging 53% protein on a dry basis. The population development strategy is detailed in Table 2. Briefly, the initial crosses were made in controlled environment growth chambers at the University of Minnesota during the winter of 1989-1990. The population was inbred via modified single seed decent (Brim, 1966) until the F_4 generation using the North American summer growing season at the University of Minnesota research farm in St. Paul, MN and a South American winter nursery location at the La Plantina Station of the Agricultural Research Institute of Chile near Santiago, Chile to gain an additional inbreeding cycle per calendar year. At harvest maturity of the F_4 population, approximately 200 lines were derived by harvesting and threshing single plants. The $F_{4:5}$ progeny rows were planted at a North American environment, and a line was selected based on high protein composition to be crossed back to the recurrent parent 'Evans'. This cycle of inbreeding, line derivation, and phenotypic evaluation of the population was carried out until the $BC_3F_{3:4}$ generation. At this point, six high-protein lines as well as three low-protein lines were identified in the subsequent progeny row yield trial in 2003 and were selected for backcrossing to 'Evans'. The purpose of divergent selection at the BC_3 was twofold: 1. To continue to develop a near isogenic population 90+% similar by pedigree to Evans that possess elevated levels of protein, and

2. To evaluate the effect a single generation of selection in the divergent direction for protein has on the resulting population (e.g. to what magnitude does protein content diminish after one selection cycle). After the Evans by BC_3:$F_{3:4}$ crosses were made, the high, and low protein populations were developed similarly to that described from initial cross to $BC_3F_{3:4}$. At the BC_4F_4 generation, 257 lines were derived from the high-protein selected populations and 135 lines were derived from the low-protein selected populations.

Field Experimentation

Plant-row experiment

Plant-row trials were carried out subsequent to line derivation in field trials in 2008. Thirty seeds from each newly derived line were planted in 1.5 M long, single-row plots at the Rosemount Research and Outreach Center at Rosemount, MN (44.7° N, 93.1° W) on a Waukegan silt loam soil type (Fine-silty over sandy or sandy-skeletal, mixed, superactive, mesic Typic Hapludoll). The experimental design was an randomized incomplete block design with 11 sets of 49 entries each. Newly derived BC_4F_4 derived lines as well as parents and checks were included in the trial. Data captured on the plant-row experiment include grain yield, maturity, and seed protein and oil content. Phenotypic values of the un-replicated genotypes were estimated after fitting an analysis of variance model that accounted for block and genotype. Block was fit as a random effect and genotype was fit as a fixed effect. The adjusted line values from the plant-row experiment were compared to the line phenotypes derived from the replicated, multi-location experiments described in the next section.

Replicated, multiple location trials

In addition to the plant-row trials, larger plot, replicated multi-location field experiments were carried out in 2010 and 2011 at the Southwest Research and Outreach Center at Lamberton, MN (44.2^{o} N, 95.3^{o} W) on a Normania loam soil (fine-loamy, mixed, superactive, mesic Aquic Hapludoll), the Rosemount Research and Outreach Center at Rosemount, MN (44.7^{o} N, 93.1^{o} W) on a Waukegan silt loam soil type (Fine-silty over sandy or sandy-skeletal, mixed, superactive, mesic Typic Hapludoll), the West Central Research and Outreach Center at Morris, MN (45.6^{o} N, 95.9^{o} W) on a Forman-Aastad complex soil type (Forman clay loam (55%) (fine-loamy, mixed, superactive, frigid Calcic Argiudoll), Aastad clay loam (20%) (fine-loamy, mixed, superactive, frigid Pachic Argiudoll), Mehurin clay loam (12%) (fine, smectitic, frigid Aquic Argiudoll), Tonka silt loam (8%) (fine, smectitic, frigid Argiaquic Argialboll), Parnell silty clay loam (3%) (fine, smectitic, frigid Vertic Argiaquolls), and Vallers loam (2%) (fine-loamy, mixed, superactive, Typic Calciaquoll)), and a farmer cooperator's field near Danvers, MN (45.3^{o} N, 95.7^{o} W) on a Bearden-Quam complex soil type (Bearden silty clay loam (60%) (fine-silty, mixed, superactive, frigid Aeric Calciaquolls), Quam silty clay loam (30%) (fine-silty, mixed, superactive, frigid Cumulic Endoaquoll), Rondell silty clay loam (7%) (fine-silty, mixed, superactive, frigid Aeric Calciaquolls), and Winger silty clay loam (3%) (fine-silty, mixed, superactive, frigid Aeric Calciaquolls).

Soybean were seeded at a density of 430,000 plants ha^{-1}. Planting dates for locations were 17 May 2010 and 17 May 2011 at Lamberton, 3 June 2010 and 9 June 2011 at Rosemount, 24 May 2010 and 25 May 2011 at Morris, and 25 May 2010 and 2

June 2011 at Danvers. Plots at Lamberton, Rosemount, and Morris were four 25-cm rows, 3.7 m long. Plots at Danvers were two 76-cm rows, 3.7 m in length. A bordered 2.4 m section in the center of the plots was harvested for seed yield determination and a subsample of approximately 0.5 kg was captured for subsequent seed composition analysis and seed size determination. Date of full maturity (R8; Fehr and Caviness, 1977) was captured on two site-years (Rosemount 2010 and Danvers 2010) and R8 canopy lodging was captured at one site-year (Rosemount 2010). The experimental design was an alpha Latinized row-column design (Williams et al., 2006) with two replications, 24 plots per column and 22 plots per row. The experimental design randomization was generated using the CycDesigN computer package (Whitaker et al., 2006). This design allowed for a total of 528 entries to be included in the trial. A total of 392 BC_4F_4 derived lines derived from the population development scheme outlined previously were included in the trial; additionally, the recurrent parent and intermediate BC parents were included. The donor parent was included in the trial; however the poor agronomic suitability precluded the capture of phenotypic data. In addition to the BC_4F_4 lines and associated parents, 130 entries from another field experiment evaluating yield and seed composition were included which reconciles the total entry number of 528 within the implemented Latinized row-column design.

Seed Composition Analysis

Cleaned soybean seed samples from each plot were subjected to near-infrared spectroscopy using a Perten DA7250 diode array instrument (Perten Instruments) with

calibration equations developed at the University of Minnesota in collaboration with Perten Instruments

Phenotypic analysis

Analysis to determine genotypic means on a site-year and combined basis was carried out for the phenotypes of protein and oil content, protein + oil content, seed yield, seed size (grams per 100 seeds; GPC), maturity and lodging. Analysis of variance models were fit for each phenotype by site-year combination using the appropriate terms based on the arrangement of the Latinized row-column experimental design (Eq.1):

$$\text{Trait} = \text{rep} + \text{column} + \text{rep*column} + \text{row [rep]} + \text{genotype} \quad \text{Eq. 1}$$

Combined analysis across site-years was carried out using the least square (LS) means calculated from the site-year specific analysis using the corresponding standard errors as weights in the analysis (Mohring and Piepho, 2009). Models for combined across site-year analysis were (Eq. 2):

$$\text{Trait} = \text{site-year} + \text{genotype} \quad \text{Eq. 2}$$

where site-year was considered a random term and genotype a fixed effect.

Trait stability across site-years evaluated was assessed via correlation analysis. Protein and yield stability was further investigated across site-years through the use of linear regression of a line's mean value at a site-year regressed against the overall site-year mean across all 392 BC_4F_4 derived lines. Similar stability analysis has been presented by previous researchers (Findlay and Wilkinson, 1963; Eberhart and Russell, 1966; Perkins and Jinks, 1968). All phenotypic data analysis was carried out with JMP 12 software (SAS Institute, Cary, NC).

Genotyping

Tissue from a random subset of 136 BC_4F_4 derived lines were collected from the first replication of the field experiment conducted at Rosemount, MN in 2010. Tissue was lyophilized and DNA was extracted using the Qiagen Biosprint 96 Plant Kit (https://www.qiagen.com/us/). The Illumina GoldenGate Assay with 1536 SNP markers were used characterize each genotype's genomic DNA (Hyten et al., 2010). SNP genotyping and allele calling was carried out by the University of Minnesota BioMedical Genomics Center.

Marker-trait association

Markers that were found to be polymorphic between 'Evans' and PI153296 based on the GoldenGate SNP data were utilized in QTL analysis. Markers that were highly heterozygous (>50%) or had a high degree of missing data (>50%) were excluded from the analysis. Likewise, genotypes that were found to be highly heterozygous or had a high rate of missing data were excluded as well. An interval mapping analysis via Haley-Knott regression (Haley and Knott, 1992) was executed through the R/qtl program (Broman et al., 2003) using 2 cM steps using the soy consensus 4.0 genetic map (Hyten et al., 2010; soybase.org). The combined phenotypes of protein, oil, yield, maturity, and protein + oil were analyzed. Highly significant, large effect QTL(s) identified in the initial genome scan were used as covariate(s) in subsequent analyses of the phenotypes indicated previously. Permutation tests with n=10000 were carried out to obtain the genome-wide LOD significance threshold for both the initial interval mapping analysis as well as the subsequent covariate QTL analysis. The protein QTL(s) identified in the

covariate analysis in addition to the QTL used as a covariate were further investigated with putative protein QTL impacts on the phenotypes of grain yield, seed size, seed oil content, the protein + oil content index, maturity, and lodging was determined by fitting the marker alleles to the phenotypic data. The QTL effect on stability across locations was assessed by evaluating the identified QTL impacts on individual site-year data.

Positioning the QTL identified in the historical context

QTL haplotype(s) identified in the present study were compared to results reported by Hwang et al. (2014), Vaughn et al. (2014), Bandillo et al. (2016). The apparent haplotype frequency of the PI 153296 high protein allele contained within the USDA germplasm collection was assessed using the SoySNP50K data on 19652 lines within the germplasm collection (Song et al., 2015; soybase.org).

Results and Discussion

Phenotypic analysis

Single site-year analysis of variance for the traits evaluated revealed significant differences ($P<0.05$) for genotypes at all locations for all traits except yield at the 2010 Danvers location (Table 2). It is evident that the spatial model experimental design aided in capturing non-genetic variation at nearly every location for each trait as indicated by the significant ($P<0.05$) effects for rows and columns. Particularly, in three of the eight site-years evaluated for yield; the rep(block) term was not significant, while row and range spatial terms were significant indicating scenarios where blocking one dimensionally is not enough to capture the spatial variability at some testing environments.

The combined analysis across environments indicated significant differences for lines ($P<0.05$) across all traits (Table 3). The random site-year effect for each trait explained more random variation than the residual term for all traits except maturity indicating that while genotype by environment interaction across traits exists to an extent, it did not explain a large portion of the variation in our data set. To further characterize the correspondence of phenotypic values obtained for the traits of interest across the different site-years, Pearson correlation coefficients were calculated for each pair-wise combination of individual site-year LS mean values. The combined analysis LS mean was included as well and pairwise combinations of individual locations by the overall analysis provides context for correlation magnitudes and aides in identifying locations that are potential anomalies relative to the other site-years. The block adjusted phenotypic

values obtained from the plant-row experiment were also included in the analysis to compare the accuracy of plant-row trials to multi-environment trial. The pairwise correlations for protein content across the data sets were generally greater than 0.9 indicating strong correspondence for the trait across site-years (Table 4). The weakest pairwise correlation between any two 2010-2011 site years was 0.865 between 2010 Morris and 2011 Rosemount and the lowest correlation between the combined analysis and any single site-year was Rosemount 2011, although the correlation was high with a value of 0.941. Ultimately, protein content is highly consistent across MN environments for the germplasm evaluated. The plant-row trial was significantly correlated with each site-year (Table 4) and the overall combined analysis ($r=0.757$, $P<0.05$). The magnitude of this plant-row to multi-location phenotype is similar to that reported by Helms and Orf (1998); while the precision of phenotypic data is diminished in a plant-row experiment compared to a multi-location trial, genetic gain for seed composition traits can be achieved using selection at this early generation. The site-year correlation comparisons for oil content (Table 5), protein + oil index (Table 6), seed size (GPC; Table 7), mirrored that of protein in that correlations between the different environments tested were high in magnitude and significance ($P<0.05$).

Pairwise Pearson correlations for yield were generally significant across all site-years tested in 2010 and 2011 with the exception of Danvers 2010 (Table 8). This location did not show a significant effect on the single location spatial analysis of variance model either, so this result was not unanticipated. Factors leading to the Danvers 2010 location not showing differences for yield could be associated with specific factors

impacting field spatial variability beyond obvious soil type or climate differences (Table 9). Soil drainage could be an explanation as the position in the field was not well drained and nearly 100 mm of rain was received from 10 through 12 August potentially creating the greater variability for yield than at other site-years (Table 8). The average pairwise correlations between all site years was approximately 0.2 indicating substantial genotype by environment interaction. The average correlation between each location and the combined yield analysis LS means was approximately 0.5 indicating that while site-year to site-year comparisons have relatively low correlations, albeit in most cases significant; the combined analysis across all locations sufficiently captures the variation exhibited across site-years. A combined analysis for yield with and without the Danvers 2011 location was compared to determine if the Danvers 2011 location should be excluded from the final, combined phenotypic analysis. In that, given the location did not show any differences for yield, the impact on the combined analysis for yield was assessed. The Pearson correlation and Spearman rank correlation coefficients between the combined yield including Danvers 2011 and the combined analysis excluding Danvers 2011 were both 0.99, indicating the location had little effect on the combined analysis. As a result, the location was removed from the combined analysis for yield. The yield data obtained for the plant-row experiment was positively correlated (0.074) with the combined analysis indicating that gain for yield can be achieved with such early generation trial similar to previous reports (Hegstad et al., 1999); although this correlation was not significant (Table 8). The three total replications of maturity data captured in 2010 as well as the plant-row data obtained in 2008 were all significantly ($P<0.05$) correlated

with one another indicating little genotype by environment interaction for maturity (Table 10).

The combined analysis LS means for each trait were compared via correlation analysis. A strong, negative correlation was observed between protein and oil ($r=-0.927$, $P<0.0001$) which is consistent with numerous previous reports (see Introduction). Of most interest is the correlation between yield and protein content which was found to have a correlation of -0.356 indicating what many previous researchers have cited as the limitation in creating lines with both high protein and yield. Oil content had a positive correlation with yield ($r=0.431$; $P<0.0001$); as did seed size ($r=0.105$; $p<0.05$). Maturity was marginally significantly associated with yield (-0.088; $P<0.10$) although the direction of the correlation is counter intuitive in that the earlier lines evaluated were slightly higher yielding than fuller-season lines. Protein content had a significant association with maturity ($r=0.409$; $p<0.001$) indicating that later maturing lines have higher protein content than earlier maturing lines which explains the negative correlation with yield and maturity; if protein is negatively associated with yield in the data set, and maturity is negatively associated with yield, then protein content would expectantly be positively correlated with maturity. Interestingly, seed size is positively correlated with protein content and yield; thus selecting for increased seed size could result in greater protein and yield content simultaneously.

Stability analysis

A scatterplot was generated (Figure 1) with yield plotted in the Y-axis and protein content plotted on the X-axis to visualize the variation for both traits across the BC_4F_4

derived population. Lines from the population that had significantly greater ($P<0.05$) overall protein content and a similar or greater line mean yield value compared to the recurrent parent 'Evans' were identified. A total of 15 lines fit this criteria and were selected for phenotypic stability characterization across the environments for which they were tested. The results of the protein stability analysis are shown in Table 12 and the yield stability results are shown in Table 13. The first analytical method of stability, regression of each lines phenotypic value at each location against the corresponding location means revealed much variation for both the regression coefficient (*b*) and r^2 with ranges from 0.38-1.32 and 0.58-0.96, respectively. The second method of phenotypic stability assessment was evaluating the ranks of the 15 genotypes and recurrent parent across the site-years sampled. Generally, lines with low regression coefficients would be considered having a stable phenotype across locations; however, lines that exhibit a low *b* will have a below average phenotype in environments that maximize the trait value. Lines that have regression coefficients near $b=1$ have a similar ranking across each environment. The variation observed in the line across environments is similar to the variation observed in general environment variation. Similarly lines with high r^2 are considered to be more stable than lines with lower r^2 values. The recurrent parent, 'Evans', was included in the stability analysis had a mean protein value of 40.03% and protein stability values of $b=1.03$ and $r^2=0.99$; thus a highly stable line with low protein content. The BC_4F_4 derived line on the other hand M04-397-6-315 is a line that has significantly greater protein content compared to 'Evans' with mean value of 43.68 and stability parameters that are similar ($b=1.13$; $r^2=0.96$; with a mean ranking across site-

years of 25. The BC_4F_4 derived line M04-403-1-1756, also with a significantly greater protein content than 'Evans' with a mean protein content of 43.24, has a *b* of 0.88 indicating that in lower protein environments it maintains its protein content better than lines with higher *b* values. To contrast, a line with among the highest mean protein content, M04-404-8-2494, is relatively unstable given the parameters for *b* (0.38) and r^2 (0.58). Evaluation of the rank stability of M04-404-8-2494 indicates that the line is among the top ten for protein content in four of the seven environments it was evaluated, but in the other environments it was ranked 24, 49 and 78. For yield, 'Evans' averaged 2.47 Mg ha^{-1} with *b*=1.39 and r^2=0.94. The BC_4F_4 derived lines that offer the greatest yield performance and stability are: M04-397-6-286, which offers better than average performance at lower yielding environments and a yield average of 2.54 Mg ha^{-1}; M04-404-11-2541, a line with a mean yield of 2.59 Mg ha^{-1} and *b*=1.01 and r^2=0.81 with an average rank across the eight site-years of 104 and demonstrating a yield advantage over 'Evans' at six of the eight site-years; and M04-404-4-2371 with a mean yield of 2.54 Mg ha^{-1}, *b*=1.06, r^2=0.79. M04-405-6-2701 is an example of a line with a high over environment mean yield, 2.59 Mg ha^{-1}, but lower stability parameters; *b*=0.46 and r^2=0.25. Considering both protein content and yield, as well as overall mean values and environmental stability, the line that best exemplifies a high protein and high yield line is M04-397-6-286. Additional lines demonstrating high protein and high yield are M04-397-6-305, M04-403-1-1756, M04-404-11-2541, M04-404-11-2553, M04-404-4-2371, and M04-405-6-2701.

QTL analysis

Interval Mapping

A total of 299 markers that were found to be of high quality and polymorphic between the donor and recurrent parent were used in analysis. A total of 119 BC_4F_4 derived lines were used in QTL mapping analysis. A visual representation of the genotypic information for the population is shown in Figure 2. The QTL analysis revealed 14 significant ($P<0.01$) QTL for protein content located on chromosome 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 16, 17, and 20 (Table 14; Figure 3). The most significant QTL was identified on chromosome 20 at the 33.2 cM position (LOD=28.12; $P<0.0001$). Two QTL were identified for yield on chromosome 14 and 20 (Table 15); the chromosome 20 region identified for yield is approximately 2 cM downstream of the protein QTL. Eleven QTL were identified for oil content on chromosomes 2, 3, 7, 9, 10, 12, 13, 14, 16, 17, 20 with the strongest QTL being the chromosome 20, 33.2 cM position which was coincident of the strongest QTL for protein content (Table 16). Eight of the 11 oil QTL were co-localized with protein QTL; the QTL on chromosome 2, 3, 10, 13, 14, 16, 17, and 20 are the same positions for both protein and oil content. Five QTL were identified for seed size (GPC; Table 17) on chromosome 6, 8, 9, 18, and 19. The QTL positions on chromosomes 8 and 9 for GPC are similar to those found for other traits. The chromosome 8 region for GPC has a similar position to a QTL identified for protein content and the chromosome 9 region found contributing to GPC was found to be similar to the chromosome 9 QTL found for protein and oil content. One QTL was identified for contributing to maturity variation located on chromosome 2 at 26.4 cM (LOD=4.73;

P=0.002); which coincides with QTL identified for protein and oil content (Table 18). A total of 14 QTL were identified for contributing to the protein + oil phenotype (P+O; Table 19). Several QTL identified for P+O are coincident with protein only QTL; the regions on chromosomes 2, 3, 4, 5, 7, 8, 9, 10, 12, 17, and 20 all map to similar positions. The P+O QTL on chromosome 6 appear to be coincident with the GPC QTL found on the same chromosome. That leaves the QTL for P+O on chromosomes 13 and 14 as potentially distinct from QTL controlling either protein or oil independently.

Interval Mapping with the chromosome 20 QTL as a covariate

A commonality of the initial QTL mapping results is the contribution of the chromosome 20, 33.2 cM position to be highly significant for controlling protein, oil, yield, and P+O phenotypes as found by several previous researchers. Given the impact of this locus on the phenotypes studied, a QTL analysis using the SNP marker at the chromosome 20, 33.2 cM position, BARC-041129-07912, was used as an additive covariate in the subsequent QTL analysis to further investigate additional QTL that contribute to the phenotypic variation. The covariate QTL analysis for protein content revealed four QTL positioned on chromosomes 4, 7, 8, and 9 (Table 20; Figure 4). The greatest LOD value came from the chromosome 4 QTL. All four of the QTL found in the covariate analysis were identified in the initial QTL analysis; nine of the QTL found initially dropped out of significance ($P<0.01$) in the covariate analysis. No QTL were identified for yield or maturity in the covariate analysis. Three QTL were found for oil content in the covariate analysis with peaks on chromosomes 2, 10, and 12 (Table 21). The regions on chromosomes 2 and 12 are coincident with QTL found in the initial QTL

analysis with QTL on chromosome 10 being unique to the covariate analysis on oil content. Six QTL were identified for GPC in the covariate analysis with five of the six being the same as those found in the initial analysis (Table 22). The single unique GPC QTL identified in the covariate analysis was positioned on chromosome 20 at 79.35 cM. A total of ten QTL for P+O were identified with the covariate analysis (Table 23); 14 were identified with the initial analysis. Nine of the ten identified in the covariate analysis coincide with QTL identified in the initial analysis with the only unique peak occurring on chromosome 20 at position 85.38 cM which is a similar position to the GPC QTL identified in the covariate analysis.

Putative protein QTL impacts and stability on protein content

Given the emphasis of the current research on increasing protein content across northerly environments while maintaining yield, we focused on the QTL identified for protein content (Table 20) including the chromosome 20 region; the impact and stability of these QTL were evaluated across the MN environments. Table 24 demonstrates the putative QTL effects at each environment. To generate this information allele calls for each marker were fit for protein content. The marker call distribution of BC_4F_4 individuals for each QTL/marker locus are included as well as the mean, standard error, upper and lower 95% confidence intervals associated with the mean allele value, the model fit R^2 as well as the significance of the model fit. Expectantly the chromosome 20 region contributed the greatest magnitude of protein content. The effects in each environment were highly significant ($P<0.05$). On average the allele contributed from the high protein donor, PI 153296, increased protein content from 2.7% at the Rosemount

environment in 2011 to 3.7% at the Morris environment in 2010; across all environments lines homozygous for the PI 153296 allele on chromosome 20 had 3.3% greater protein content (Table 25). The favorable alleles of the four other QTL identified as likely contributing to protein content were contributed from the recurrent parent Evans. The QTL on chromosome 4 had an across all site-year effect of 2.3%, the chromosome 7 region had an average impact of 2.1%, the QTL on chromosome 8 had an overall effect of 2.0%, and the QTL positioned on chromosome 9 had an effect of 2.2% (Table 25). The frequency of the marker alleles among the BC_4F_4 lines for the four QTL mentioned is highly skewed toward the Evans inherited allele indicating that the phenotypic selection and backcrossing throughout the development of the population enriched the frequency of the favorable allele from the recurrent parent. Interestingly, the protein QTL had similar impact on the PYT data indicating that smaller, replicated plots generated similar information for protein content as did multi-location, replicated trials.

Putative protein QTL impacts and stability on grain yield

Given the reported phenotypic analysis indicating a negative yield: maturity correlation, one would logically assume that the QTL which increase protein content decrease yield. The protein QTL yield characterization information is given in Table 26, and this trend, at least numerically, is consistent. There are environments were the yield impact is not significantly different for the high protein allele such as Lamberton for both 2010 and 2011 for the chromosome 20 region (Table 27). However, the other six individual environments as well as the overall phenotypic analysis indicates that there is negative yield associated with the increased protein content in the context of the

chromosome 20 QTL. The direction of the yield effect for the chromosome 4, 7, 8, and 9 regions is similarly consistent in direction, however, generally the high protein allele does not cause significant ($P<0.05$) yield reductions (Table 27). The putative protein QTL on chromosomes 4, 7, 8, and 9 are intriguing in that they appear to be contributing the elevated protein content while not significantly reducing yield; however, pragmatically, these favorable protein QTL already exist in the elite MG1 breeding pool thus do not represent a great breeding opportunity unless an elite x elite population is segregating for them. The protein QTL with greatest potential to elevate seed protein content is the large effect QTL on chromosome 20. As indicated in the introduction, this not the first instance of identification of the chromosome 20 region, however it is the first occurrence of the QTL confirmed in the high protein accession PI1 53296, and characterized across MN environments where elevated protein content offers value to soybean processors and producers. Similar to Figure 1, Figure 5 displays grain yield regressed against seed protein content for BC_4F_4 derived lines genotyped and included in the QTL analysis with indicators for chromosome 20 parental contributions.

Comparison of current results to previous findings

The chromosome 20 region identified and characterized in the current work has been well documented (cqSeed protein-003; Nichols et al., 2006) via family based QTL mapping and genome wide association analysis. Given the recent work by Hwang et al. (2014), Vaughn et al. (2014), and Bandillo et al. (2015), the genomic construction (e.g. haplotypes, gene models, etc.) of the region, is more defined. The current study, while focusing on the agronomic attributes of the region for breeding utility in the context of a

bi-parental, advanced-backcross population, the data set is limited for robust *de novo* characterization of the genomic region. However, given both parents in the current study were assayed with the Soy50K SNP assay, we can make comparisons between the parents used in the current study and the regions defined by Hwang et al. (2014), Vaughn et al. (2014), and Bandillo et al. (2015).

Interestingly, each marker identified by Hwang et al. (2014), Vaughn et al. (2014), and Bandillo et al. (2015) as being significant are monomorphic between 'Evans' and PI 153296. The physical position of the BARC-041129-07912 identified as the most significant in the current study is 32,449,414 bp (G.max assembly 1.01, same map used in the previous studies) which is slightly downstream of the region identified by Vaughn et al. (2014) and Bandillo et al. (2015) which ends at the 32,052,917 bp position. Likely the most probable, most significant, diagnostic markers for this region are six markers on the Soy50KSNP panel that are polymorphic between 'Evans' and PI 15296 located from 30,564,816-30,837,430 bp which is in haplotype block four identified by Bandillo et al. (2016). Alternatively, there is another polymorphic block around 32.6 Mbp which includes the GoldenGate SNP found to be most significant in the present study. This region was identified as haplotype block 5 by Bandillo et al., (2015) (Figure 6.). While it has been identified that the historical data within GRIN may not be the most optimum to validate/assess marker effects due to the difference in the temporal and geographic distribution of the field phenotyping trials, it is the most complete publically available dataset that exists for protein content across the set of accessions. The lines within the germplasm collection that had identical marker calls to PI 153296 and Evans were

compared for both haplotype block regions delineated by Bandillo et al. (2015) for seed protein content data contained within GRIN.

Of the 19651 lines assessed for SoySNP50K data, a total of 14863 lines have protein data in the GRIN database. In the haplotype block 4 region, 29 lines have the same marker haplotype as PI 153296 and 14757 had the same marker haplotype as Evans. For the block 5 region there were 416 lines had the same marker haplotype as PI 153296 whereas 7540 lines had the same marker haplotype as Evans. For the haplotype 4 region, PI 153296 like lines had a mean protein content of 49.5% where the Evans like lines had a protein content of 44.2% (Table 27). For the haplotype 5 region. PI 153296 like lines had a protein content of 45.7% and Evans like lines had a mean protein content of 43.8%. These differences when subjected to a t-test show a significant difference between the PI 153296 and Evans haplotypes for the two regions on chromosome 20 (Table 27).

Conclusions

Through four cycles of crossing, inbreeding, phenotyping, and selection; seed protein content was successfully increased within the recurrent parent 'Evans'. While protein was increased, yield was conserved at a level where yield was not significantly different than the recurrent parent, resulting in the creation of many high-protein, high-yielding lines. This increase did come at the expense of oil content, however, the geography for which the germpasm is adapted, the protein increase at the expense of oil results in a likely net gain in value to the soybean crop. The genetic architecture contributing to the elevated protein was found to be similar to that of previous reports; the major QTL on chromosome 20 was the primary driver for elevated protein. While this QTL has been identified by several previous investigators, this work is the first empirical information indicating the elevated protein found in PI 153296 is attributable to this chromosome 20 region. Enabled through the genetic characterization of the USDA-GRIN soybean germplasm with the SoySNP50K assay, and subsequent association analysis, comparisons within the segregating regions between the PI 153296 high protein donor and Evans can be made. It was found that that Evans and PI 153296 are polymorphic at two haplotype blocks described by Bandillo et al. (2015). Bolon et al. (2010) in the initial genomic assessment of the chromosome 20 region identified 12 potential candidate genes. Of these 12, Bandillo et al. (2015) identified three that were the most probable given previous linkage and association mapping experiments as well as their own association analysis. Two of the three genes they found to be the most probable exist within each of the two haplotype blocks for which Evans and PI 153296 are polymorphic.

Glyma20g21361 in haplotype block 4 which is associated with a conserved oligomeric Golgi complex (subunit 6) which is involved in the transfer and storage of proteins throughout the intra- and intercellular vesicular-space and Glyma20G21780 contained within haplotype block 5 which encodes an ethylene receptor shown to be associated with signal transduction and protein histidine kinase activity (Bolon et al., 2010). The third most probable gene model is located in a region not segregating between Evans and PI 153296. Given the relative sparsity of the marker density used in the present work, only a single marker was polymorphic in the vicinity, thus higher density evaluation of the genetic structure of this region among the advanced backcross progeny could not be executed. The frequency of 'Evans' vs PI 153296 haplotypes in the two blocks were assessed and a comparison of mean protein content for the haplotypes revealed a low frequency of the high protein haplotype (based on the PI 153296 line).

Given the observed impact the chromosome 20 protein QTL region has on seed protein seed composition, soybean breeders may consider further utilizing the QTL to increase protein content in breeding lines. While the often reported negative association between yield and protein was observed in the present study, it appears as though higher seed protein content lines may be created and identified that do not experience a precipitous loss in grain yield.

Table 1. Population development strategy and timeline for the advanced backcross population between Evans (recurrent parent) and PI153296 (high protein donor).

Location	N.A. Season	Task	Timeframe	Generation
St. Paul, MN	Winter	Crossing	1989-1990	RP x Donor
St. Paul, MN	Summer	F1 growout	1990	F_1
Santiago, Chile	Winter	F2 growout	1990-1991	F_2
St. Paul, MN	Summer	F3 growout	1991	F_3
Santiago, Chile	Winter	F4 growout, Line derivation	1991-1992	F_4
St. Paul, MN	Summer	Progeny row trial	1992	$F_{4:5}$
St. Paul, MN	Summer	Crossing	1993	RP x $F_{4:5}$
Santiago, Chile	Winter	F1 growout	1993-1994	BC_1F_1
St. Paul, MN	Summer	F2 growout	1994	BC_1F_2
Santiago, Chile	Winter	F3 growout	1994-1995	BC_1F_3
St. Paul, MN	Summer	F4 growout, Line derivation	1995	BC_1F_4
St. Paul, MN	Summer	Progeny row trial	1996	$BC_1F_{4:5}$
St. Paul, MN	Summer	Crossing	1997	RP x $BC_1F_{4:5}$
Santiago, Chile	Winter	F1 growout	1997-1998	BC_2F_1
St. Paul, MN	Summer	F2 growout	1998	BC_2F_2
Santiago, Chile	Winter	F3 growout	1998-1999	BC_2F_3
St. Paul, MN	Summer	F4 growout, Line derivation	1999	BC_2F_4
St. Paul, MN	Summer	Progeny row trial	2000	$BC_2F_{4:5}$
St. Paul, MN	Summer	Crossing	2001	RP x $BC_2F_{4:5}$
Santiago, Chile	Winter	F1 growout	2001-2002	BC_3F_1
St. Paul, MN	Summer	F2 growout	2002	BC_3F_2
Santiago, Chile	Winter	F3 growout, Line derivation	2002-2003	BC_3F_3
St. Paul, MN	Summer	Progeny row trial	2003	$BC_3F_{3:4}$
St. Paul, MN	Summer	Crossing	2004	RP x $BC_3F_{3:4}$
Santiago, Chile	Winter	F1 growout	2004-2005	BC_4F_1
St. Paul, MN	Summer	F2 growout	2005	BC_4F_2
Santiago, Chile	Winter	F3 growout	2005-2006	BC_4F_3
St. Paul, MN	Summer	F4 growout, Line derivation	2006	BC_4F_4
St. Paul, MN	Summer	Progeny row trial	2008	$BC_4F_{4:5}$

Table 2. Analysis of variance for seed protein and oil concentration, seed yield, seed size, maturity, and lodging for lines grown at eight Minnesota site-years in a Latinized row-column design.

		Danvers						Lamberton						Morris						Rosemount					
		2010			2011			2010			2011			2010			2011			2010			2011		
Source of variation	Trait	df	MS		df	MS		df	MS		df	MS		df	MS		df	MS		df	MS		df	MS	
Row[Rep]	Protein	42	0.7	†	42	0.4	NS	42	0.3	NS	----	----	---	42	1.2	†	42	0.9	†	42	0.6	†	42	0.9	†
Line	Protein	523	6.5	†	523	6.2	†	523	6.7	†	----	----	---	523	7.8	†	523	5.8	†	523	5.5	†	523	4.3	†
Rep	Protein	1	0.1	NS	1	14.8	†	1	6.7	†	----	----	---	1	0.1	NS	1	16.3	†	1	8.5	†	1	39.2	†
Rng	Protein	23	1.0	†	23	0.7	*	23	0.5	*	----	----	---	23	0.9	*	23	7.1	†	23	2.3	†	23	3.0	†
Rng*Rep	Protein	23	0.9	†	23	0.6	*	23	0.3	NS	----	----	---	23	0.5	NS	23	1.1	†	23	0.4	*	23	1.7	†
Error	Protein	434	0.2		434	0.4		434	0.3		----	----	---	434	0.4		434	0.3		434	0.2		434	0.4	
Row[Rep]	Oil	42	0.3	†	42	0.3	*	42	0.2	NS	----	----	---	42	0.5	†	42	0.3	†	42	0.2	*	42	1.2	†
Line	Oil	523	2.4	†	523	1.5	†	523	2.6	†	----	----	---	523	3.0	†	523	1.3	†	523	2.1	†	523	1.2	†
Rep	Oil	1	31.9	†	1	0.0	NS	1	17.5	†	----	----	---	1	4.4	†	1	14.9	†	1	4.2	†	1	14.3	†
Rng	Oil	23	0.7	†	23	0.2	*	23	0.2	*	----	----	---	23	0.8	†	23	2.1	†	23	0.3	†	23	0.7	†
Rng*Rep	Oil	23	0.3	*	23	0.2	*	23	0.4	†	----	----	---	23	0.3	NS	23	0.3	†	23	0.2	*	23	0.3	*
Error	Oil	434	0.1		434	0.2		434	0.1		----	----	---	434	0.2		434	0.1		434	0.1		434	0.1	
Row[Rep]	Yield	42	0.1	NS	42	0.8	†	42	0.3	†	42	0.5	†	42	0.6	†	41	0.4	†	42	0.2	†	42	0.4	†
Line	Yield	524	0.1	NS	522	0.2	†	521	0.2	†	536	0.6	†	522	0.1	†	533	0.3	†	522	0.1	†	533	0.1	†
Rep	Yield	1	0.0	NS	1	6.3	†	1	0.0	NS	1	0.5	NS	1	1.0	*	1	15.5	†	1	4.7	†	1	5.1	†
Rng	Yield	23	0.4	*	23	0.3	†	23	0.5	†	23	3.1	†	23	1.2	†	23	1.7	†	23	0.1	†	23	0.1	†
Rng*Rep	Yield	23	0.3	*	23	0.2	†	23	0.3	†	23	0.4	*	23	0.5	†	23	0.7	†	23	0.1	*	23	0.1	NS
Error	Yield	431	0.1		430	0.1		415	0.1		428	0.2		425	0.1		425	0.1		427	0.0		414	0.0	
Row[Rep]	GPC	42	1.5	†	42	2.4	†	42	0.3	NS	42	0.6	*	42	1.7	†	41	4.2	†	42	0.3	*	42	0.6	†
Line	GPC	523	1.1	†	522	1.3	†	521	1.5	†	536	1.3	†	522	1.4	†	532	1.5	†	522	1.9	†	532	1.4	†
Rep	GPC	1	35.8	†	1	234.0	†	1	0.5	NS	1	37.2	†	1	18.7	†	1	0.3	NS	1	0.3	NS	1	127.5	†
Rng	GPC	23	1.1	†	23	1.3	†	23	2.0	†	23	1.4	†	23	1.4	†	23	1.0	*	23	0.8	†	23	1.1	†
Rng*Rep	GPC	23	0.7	†	23	0.5	*	23	0.7	†	23	0.8	†	23	0.8	†	23	0.8	*	23	0.5	†	23	0.5	*
Error	GPC	427	0.2		418	0.3		419	0.2		416	0.3		430	0.3		412	0.5		426	0.2		401	0.3	
Row[Rep]	Maturity	‡	‡	‡	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	42	6.9	†	----	----	---
Line	Maturity	‡	‡	‡	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	523	18.9	†	----	----	---
Rep	Maturity	‡	‡	‡	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	1	40.9	*	----	----	---
Rng	Maturity	‡	‡	‡	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	23	13.9	†	----	----	---
Rng*Rep	Maturity	‡	‡	‡	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	23	5.8	*	----	----	---
Error	Maturity	‡	‡	‡	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	428	2.8		----	----	---
Row[Rep]	Lodging	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	42	0.2	NS	----	----	---
Line	Lodging	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	522	0.4	†	----	----	---

Rep	Lodging	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	1	0.5	NS	----	----	---
Rng	Lodging	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	23	0.4	*	----	----	---
Rng*Rep	Lodging	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	23	0.2	NS	----	----	---
Error	Lodging	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	----	----	---	428	0.2		----	----	---

*Significant at P<0.05

†Significant at P<0.0001

‡Single rep of data captured

Table 3. Combined across site-years mixed model analysis summary for soybean traits captured across eight Minnesota site-years in 2010 and 2011.

	Random effect % variation		Fixed effect significance
Trait	Site-year	Residual	Line
Protein†	87.4	12.6	***
Yield‡	76.1	23.9	***
Oil†	79.6	20.4	***
P+O†	59.2	40.8	***
GPC†	84.4	15.6	***
MAT§	13.1	86.9	***

***Significant at P<0.0001

†Data represents a total of 7 MN site-years

‡Data represents a total of 8 MN site-years

§Data represents a total of 3 reps across 2 MN site-years

Table 4. Pearson correlation coefficients and corresponding significance indicators for soybean seed protein concentration for BC_4F_4-derived lines grown at different Minnesota site-years.

	PRO ALL	PRO-Danv2010	PRO-Danv2011	PRO-Lamb2010	PRO-Morr2010	PRO-Morr2011	PRO-Rose2010	PRO-Rose2011	PRO-PYT
PRO ALL	----	***	***	***	***	***	***	***	***
PRO-Danv2010	0.964	----	***	***	***	***	***	***	***
PRO-Danv2011	0.963	0.914	----	***	***	***	***	***	***
PRO-Lamb2010	0.965	0.934	0.908	----	***	***	***	***	***
PRO-Morr2010	0.958	0.918	0.898	0.914	----	***	***	***	***
PRO-Morr2011	0.967	0.924	0.928	0.918	0.908	----	***	***	***
PRO-Rose2010	0.952	0.904	0.905	0.917	0.895	0.913	----	***	***
PRO-Rose2011	0.941	0.885	0.904	0.889	0.865	0.905	0.877	----	***
PRO-PYT	0.757	0.731	0.728	0.724	0.724	0.742	0.732	0.699	----

***Significant at P<0.0001

Table 5. **Pearson correlation coefficients and corresponding significance indicators for soybean seed oil concentration for BC_4F_4-derived lines grown at different Minnesota site-years.**

	Oil ALL	OIL-Danv2010	OIL-Danv2011	OIL-Lamb2010	OIL-Morr2010	OIL-Morr2011	OIL-Rose2010	OIL-Rose2011	OIL-PYT
Oil ALL	----	***	***	***	***	***	***	***	***
OIL-Danv2010	0.956	----	***	***	***	***	***	***	***
OIL-Danv2011	0.936	0.876	----	***	***	***	***	***	***
OIL-Lamb2010	0.954	0.907	0.860	----	***	***	***	***	***
OIL-Morr2010	0.953	0.906	0.868	0.892	----	***	***	***	***
OIL-Morr2011	0.913	0.849	0.848	0.852	0.815	----	***	***	***
OIL-Rose2010	0.960	0.902	0.884	0.912	0.907	0.857	----	***	***
OIL-Rose2011	0.895	0.815	0.830	0.831	0.790	0.854	0.848	----	***
OIL-PRT	0.770	0.744	0.717	0.736	0.727	0.709	0.729	0.703	----

***Significant at P<0.0001

Table 6. **Pearson correlation coefficients and corresponding significance indicators for the index (sum) of soybean seed protein and oil concentration for BC_4F_4-derived lines grown at different Minnesota site-years.**

	P+O ALL	P+O-Danv2010	P+O-Danv2011	P+O-Lamb2010	P+O-Morr2010	P+O-Morr2011	P+O-Rose2010	P+O-Rose2011	P+O-PYT
P+O ALL	----	***	***	***	***	***	***	***	***
P+O-Danv2010	0.870	----	***	***	***	***	***	***	***
P+O-Danv2011	0.882	0.822	----	***	***	***	***	***	***
P+O-Lamb2010	0.931	0.714	0.700	----	***	***	***	***	***
P+O-Morr2010	0.879	0.830	0.841	0.707	----	***	***	***	***
P+O-Morr2011	0.860	0.778	0.865	0.676	0.824	----	***	***	***
P+O-Rose2010	0.845	0.796	0.814	0.693	0.795	0.797	----	***	***
P+O-Rose2011	0.829	0.774	0.846	0.627	0.811	0.819	0.752	----	***
P+O-PRT	0.621	0.555	0.618	0.520	0.579	0.632	0.591	0.556	----

***Significant at P<0.0001

Table 7. Pearson correlation coefficients and corresponding significance indicators for soybean seed size (grams per 100 seeds; GPC) for BC_4F_4-derived lines grown at different Minnesota site-years.

	GPC ALL	GPC-Danv2010	GPC-Danv2011	GPC-Lamb2010	GPC-Lamb2011	GPC-Morr2010	GPC-Morr2011	GPC-Rose2010	GPC-Rose2011
GPC ALL	----	***	***	***	***	***	***	***	***
GPC-Danv2010	0.812	----	***	***	***	***	***	***	***
GPC-Danv2011	0.831	0.640	----	***	***	***	***	***	***
GPC-Lamb2010	0.707	0.549	0.559	----	***	***	***	***	***
GPC-Lamb2011	0.765	0.554	0.562	0.536	----	***	***	***	***
GPC-Morr2010	0.826	0.649	0.614	0.536	0.578	----	***	***	***
GPC-Morr2011	0.800	0.557	0.606	0.466	0.539	0.564	----	***	***
GPC-Rose2010	0.801	0.725	0.613	0.369	0.595	0.725	0.530	----	***
GPC-Rose2011	0.842	0.632	0.719	0.578	0.561	0.639	0.629	0.640	----

***Significant at P<0.0001

Table 8. Pearson correlation coefficients and corresponding significance indicators for soybean seed yield for BC_4F_4-derived lines grown at different Minnesota site-years.

	Yield ALL	Yield-Danv2010	Yield-Danv2011	Yield-Lamb2010	Yield-Lamb2011	Yield-Morr2010	Yield-Morr2011	Yield-Rose2010	Yield-Rose2011	Yield-PYT
Yield ALL	----	***	***	***	***	***	***	***	***	NS
Yield-Danv2010	0.284	----	NS	NS	NS	NS	NS	NS	NS	NS
Yield-Danv2011	0.530	0.025	----	*	***	*	***	*	***	*
Yield-Lamb2010	0.552	0.049	0.159	----	***	***	*	***	†	NS
Yield-Lamb2011	0.704	0.023	0.166	0.305	----	*	†	*	NS	NS
Yield-Morr2010	0.393	-0.014	0.102	0.247	0.140	----	NS	***	***	NS
Yield-Morr2011	0.559	0.036	0.437	0.115	0.093	0.056	----	*	***	†
Yield-Rose2010	0.355	-0.019	0.158	0.289	0.139	0.289	0.119	----	NS	NS
Yield-Rose2011	0.456	0.058	0.448	0.098	0.044	0.165	0.498	0.068	----	NS
Yield-PRT	0.074	0.051	0.109	0.050	0.005	-0.008	0.097	0.013	0.027	----

***Significant at P<0.0001
*Significant at P<0.05
†Significant at P<0.10
NS-not significant

Table 9. Monthly weather information for environments where BC_4F_4 lines were characterized including average mean daily temperature (degrees C), precipitation (mm), and growing degree days (GDD; base 10).

		Mean Temp (C)	Precip (mm)	GDD_{10}	Mean Temp. (C)	Precip (mm)	GDD_{10}
Location	Month		2010			2011	
Danvers	May	14.2	20.1	187	13.1	70.4	149
	June	19.0	84.3	262	19.1	65.3	273
	July	22.1	44.5	353	23.9	94.7	429
	August	22.0	124.7	353	20.4	24.6	310
	September	14.1	53.8	148	14.8	8.1	165
	October	9.5	8.1	104	10.4	25.1	136
	Avg.	16.8	55.9	235	17.0	48.0	244
	Sum	----	391.5	1642	----	336.3	1708
Morris	May	13.8	63.5	244	12.1	146.1	202
	June	19.0	85.1	349	18.3	71.1	321
	July	21.8	79.8	440	23.6	188.0	487
	August	21.8	201.4	444	20.1	64.0	386
	September	13.4	130.3	218	14.7	11.4	265
	October	9.3	53.6	182	10.1	23.4	188
	Avg.	16.5	102.3	313	16.5	84.0	308
	Sum	----	715.9	2189	----	587.9	2156
Rosemount	May	15.9	56.4	218	14.4	85.1	180
	June	20.2	115.3	319	20.2	75.2	311
	July	23.8	46.2	435	25.1	108.7	472
	August	24.4	73.7	449	22.3	90.2	388
	September	15.5	83.3	190	16.5	10.9	215
	October	11.9	34.3	125	12.6	9.9	154
	Avg.	18.6	68.2	289	18.5	63.3	287
	Sum	----	477.4	2025	----	443.3	2007
Lamberton	May	14.5	26.4	190	12.8	53.1	146
	June	19.4	160.8	290	19.5	181.4	292
	July	22.3	24.4	384	24.3	58.4	447
	August	22.8	77.0	397	20.8	28.4	327
	September	14.9	239.0	166	15.4	2.0	179
	October	11.3	2.0	127	11.7	9.4	146
	Avg.	17.5	88.3	259	17.4	55.5	256
	Sum	----	617.9	1814	----	388.2	1794

Table 10. Pearson correlation coefficients and corresponding significance indicators for soybean maturity (MAT) for BC_4F_4-derived lines grown at different Minnesota rep-site-years.

	MAT ALL	MAT-Danv	MAT-Rose 2	MAT-Rose1	MAT-PYT
MAT ALL	----	***	***	***	***
MAT-Danv	0.840	----	***	***	***
MAT-Rose 2	0.852	0.541	----	***	***
MAT-Rose1	0.843	0.528	0.653	----	***
MAT-PRT	0.665	0.528	0.578	0.586	----

***Significant at P<0.0001

Table 11. Pearson correlation coefficients and corresponding significance indicators for the pairwise comparisons of different soybean traits. Phenotypic data used for each trait represents the combined analysis LS means across field experiments conducted across Minnesota site-years in 2010 and 2011.

	PRO ALL	Oil ALL	Yield ALL	GPC ALL	LDG	MAT ALL	P+O ALL
PRO ALL	----	***	***	***	***	***	***
Oil ALL	-0.927	----	***	***	***	***	***
Yield ALL	-0.356	0.431	----	*	NS	†	***
GPC ALL	0.295	-0.147	0.105	----	NS	*	***
LDG	0.223	-0.263	-0.056	0.079	----	***	*
MAT ALL	0.409	-0.521	-0.088	0.138	0.450	----	***
P+O ALL	0.856	-0.640	-0.180	0.383	0.128	0.205	----

***Significant at P<0.0001
*Significant at P<0.05
†Significant at P<0.10

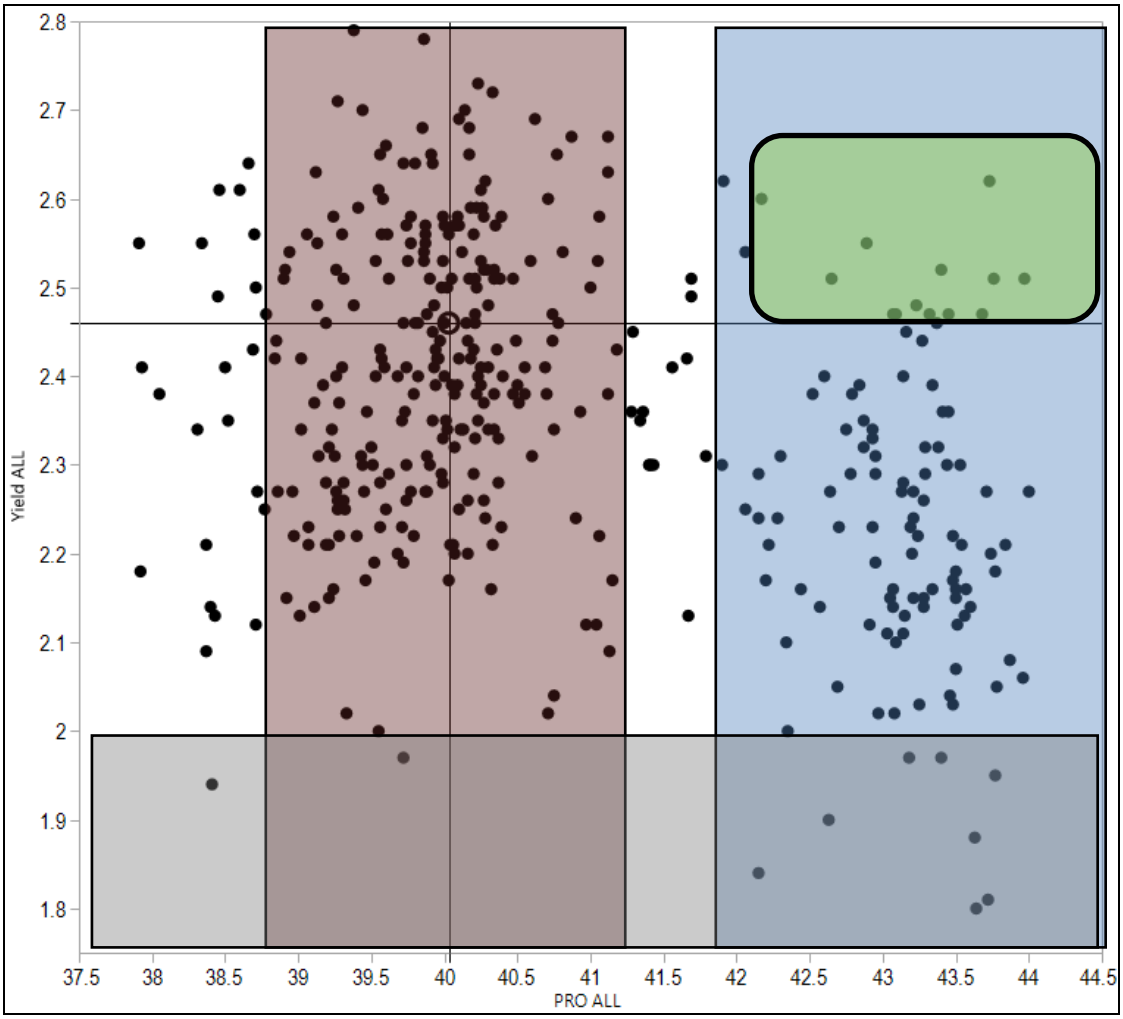


Figure 1. Combined least square means for seed yield (Mg ha^{-1}) vs seed protein concentration (% dry basis) for BC_4F_4 derived lines developed from the recurrent parent 'Evans' and the high protein donor parent 'PI153296'. The solid lines originating on the Y-axis (2.36) and X-axis (40.03) denote the mean value for yield and protein of the recurrent parent. The red shaded box indicates the 95% confidence interval around the mean protein concentration of 'Evans'. The blue shaded box denotes the lines that possess significantly (alpha=0.05) greater protein than 'Evans'. The gray shaded box indicates the lines that have significantly lower yield than 'Evans'. The green shaded box identifies 15 BC_4F_4 derived lines that have apparent greater protein and, at least, equivalent yield compared to the recurrent parent 'Evans'.

Table 12. Subset results of stability analysis for protein content. Analysis was conducted across the set of all 392 BC_4F_4 derived lines, with the results in the table representing the 15 BC_4F_4 derived lines that were found to have greater protein content and similar to greater yield than the recurrent parent 'Evans'. Regression analysis shows the coefficients associated with model fits of each line's environmental mean value regressed against the location mean value. The phenotypic rankings of the 15 lines and the recurrent parent relative to all 392 lines are shown.

	Protein													
	Regression analysis					Rank comparisons								
Line	Mean	95% CI	b	r2	P-value indicator	Danv2010	Danv2011	Lamb2010	Morr2010	Morr2011	Rose2010	Rose2011	Overall mean	Mean rank across 7 environments
	% Protein													
EVANS	40.03	1.18	1.03	0.99	**	225	206	215	237	240	236	268	230	232.4
M04-397-6-286	43.46	1.07	1.32	0.87	**	32	54	4	92	23	12	49	35	38.0
M04-397-6-305	43.97	1.08	0.86	0.70	*	4	3	2	72	11	31	1	2	17.7
M04-397-6-311	42.07	1.08	0.99	0.73	*	60	107	97	107	104	125	104	105	100.6
M04-397-6-315	43.68	1.08	1.13	0.96	**	7	23	22	38	57	17	9	14	24.7
M04-403-1-1756	43.24	1.08	0.88	0.89	**	29	26	60	76	52	63	30	50	48.0
M04-403-1-1772	43.32	1.08	0.89	0.84	**	31	72	68	29	27	51	36	41	44.9
M04-403-13-2159	43.08	1.08	0.91	0.66	*	36	27	36	105	59	85	20	68	52.6
M04-403-16-2241	43.44	1.08	1.12	0.81	**	20	32	48	17	87	28	45	32	39.6
M04-403-16-2246	43.07	1.07	0.93	0.75	**	100	45	25	42	79	66	48	63	57.9
M04-403-9-2036	43.38	1.08	1.08	0.87	**	45	71	53	27	31	16	54	38	42.4
M04-404-11-2541	42.15	1.08	0.89	0.89	**	99	102	98	102	106	115	81	101	100.4
M04-404-11-2553	42.70	1.08	1.24	0.83	**	73	89	63	90	75	38	109	88	76.7
M04-404-4-2371	42.88	1.08	0.71	0.76	**	93	87	65	81	20	90	29	79	66.4
M04-404-8-2494	43.91	1.16	0.38	0.58	t	24	8	78	7	---	49	7	9	28.8
M04-405-6-2701	43.74	1.08	1.10	0.81	**	23	31	47	3	26	3	39	11	24.6

**Significant at P<0.01

* Significant at P<0.05

†Significant at P<0.10

Table 13. Subset results of stability analysis for grain yield. Analysis was conducted across the set of all 392 BC_4F_4 derived lines, with the results in the table representing the 15 BC_4F_4 derived lines that were found to have greater protein content and similar to greater yield than the recurrent parent 'Evans'. Regression analysis shows the coefficients associated with model fits of each line's environmental mean value regressed against the location mean value. The phenotypic rankings of the 15 lines and the recurrent parent relative to all 392 lines are shown.

	Yield														
	Regression analysis					Rank comparisons									
Line	Mean	95% C.I.	b	r^2	P-value indicator	Danv2010	Lamb2010	Lamb2011	Morr2010	Morr2011	Rose2010	Rose2011	Overall mean	Mean rank across 7 environments
	Mg ha^{-1}														
EVANS	2.47	0.45	1.39	0.94	**	236	41	156	199	211	149	188	121	162.6
M04-397-6-286	2.54	0.37	0.74	0.52	*	96	43	124	51	229	7	224	80	106.8
M04-397-6-305	2.47	0.38	0.99	0.47	t	147	32	44	127	259	193	326	87	151.9
M04-397-6-311	2.47	0.38	0.86	0.24	NS	182	303	8	136	23	342	303	69	170.8
M04-397-6-315	2.42	0.38	1.70	0.52	*	374	20	6	355	311	315	310	119	226.3
M04-403-1-1756	2.47	0.38	1.02	0.58	*	244	253	51	153	263	111	219	106	175.0
M04-403-1-1772	2.43	0.38	1.26	0.72	**	50	54	107	301	139	345	246	115	169.6
M04-403-13-2159	2.45	0.38	0.89	0.58	*	179	37	129	31	238	210	294	118	154.5
M04-403-16-2241	2.45	0.38	1.15	0.56	*	364	114	42	215	123	83	203	114	157.3
M04-403-16-2246	2.44	0.37	1.18	0.52	*	286	268	22	344	160	250	148	113	198.9
M04-403-9-2036	2.38	0.38	0.43	0.04	NS	9	107	7	251	381	129	389	123	174.5
M04-404-11-2541	2.59	0.38	1.01	0.81	**	189	62	77	44	92	108	108	33	89.1
M04-404-11-2553	2.48	0.38	0.81	0.39	t	37	347	55	279	114	274	62	89	157.1
M04-404-4-2371	2.54	0.38	1.06	0.79	**	46	194	166	128	30	59	195	66	110.5
M04-404-8-2494	2.42	0.41	1.34	0.61	*	159	116	46	206	---	136	378	88	161.3
M04-405-6-2701	2.59	0.38	0.46	0.25	NS	52	265	26	74	99	133	19	27	86.9

**Significant at P<0.01

* Significant at P<0.05

†Significant at P<0.10

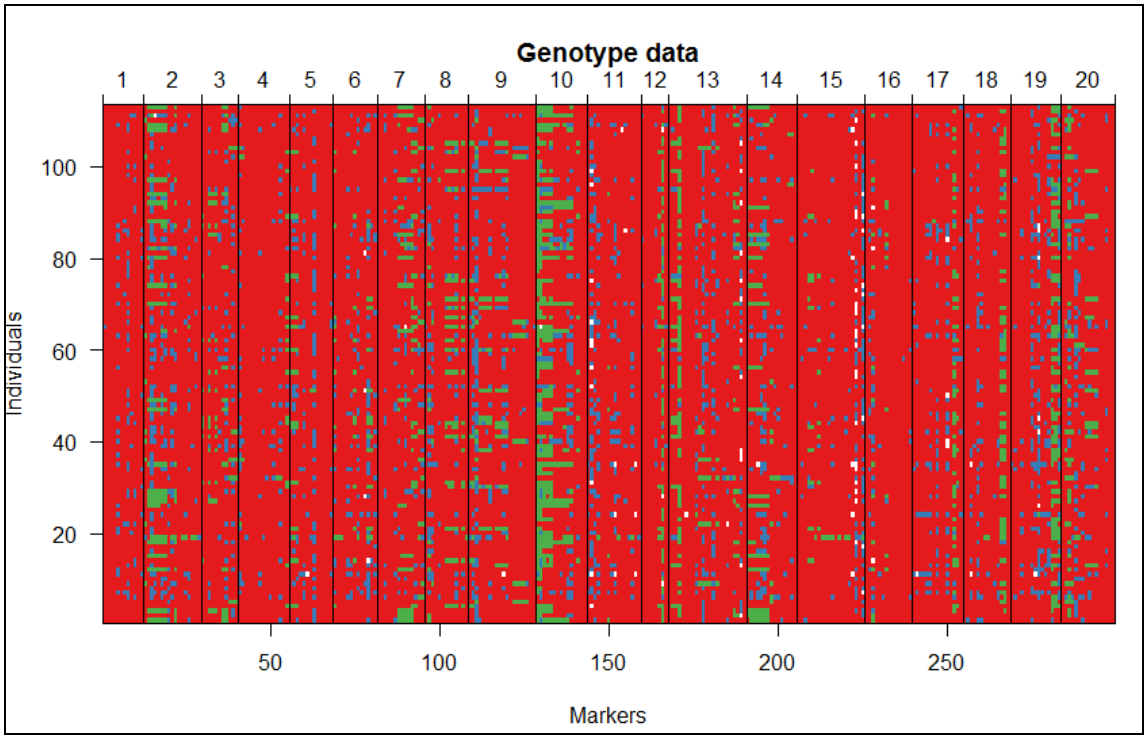


Figure 2. Line by SNP marker origin heatmap; red indicates "Evans" origin, green indicates PI153296 origin. Blue points are heterozygous marker calls and white indicates missing marker calls.

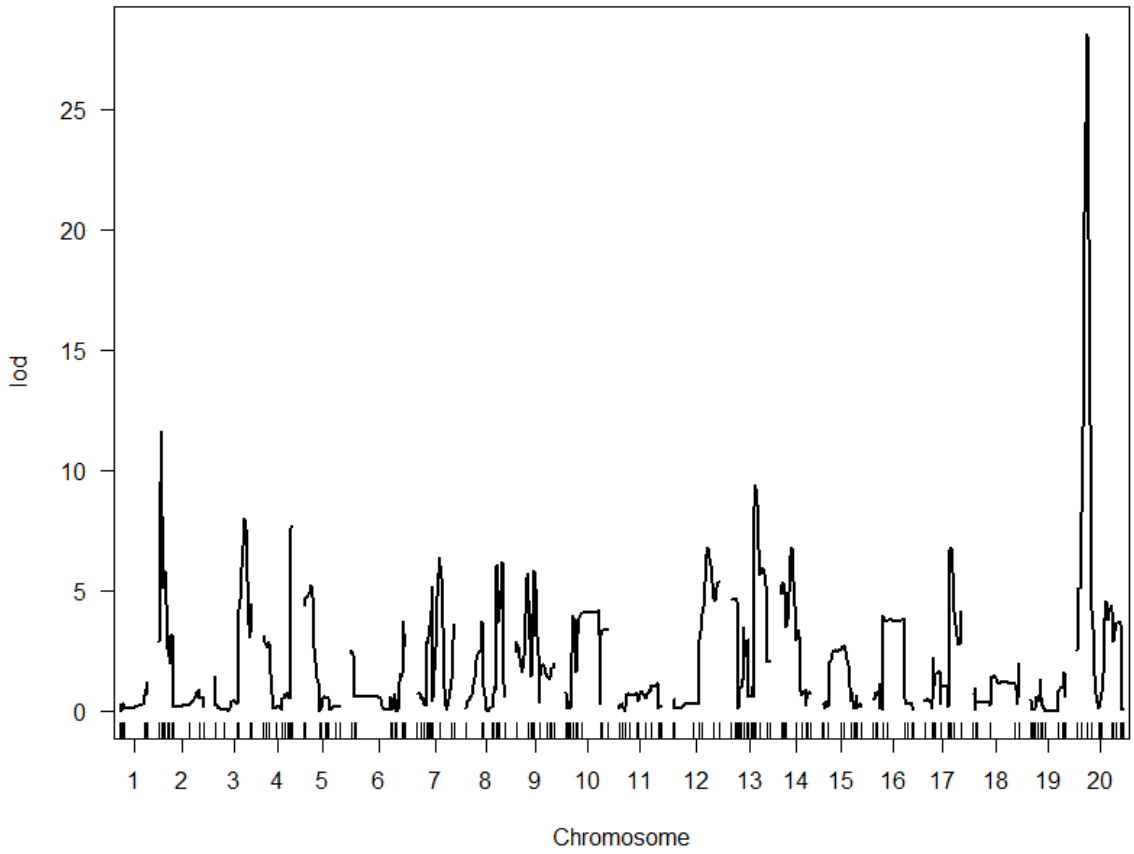


Figure 3. Manhattan plot showing the logarithm of odds (LOD) across soybean genome positions from the interval mapping QTL analysis for protein content of BC_4F_4 derived lines.

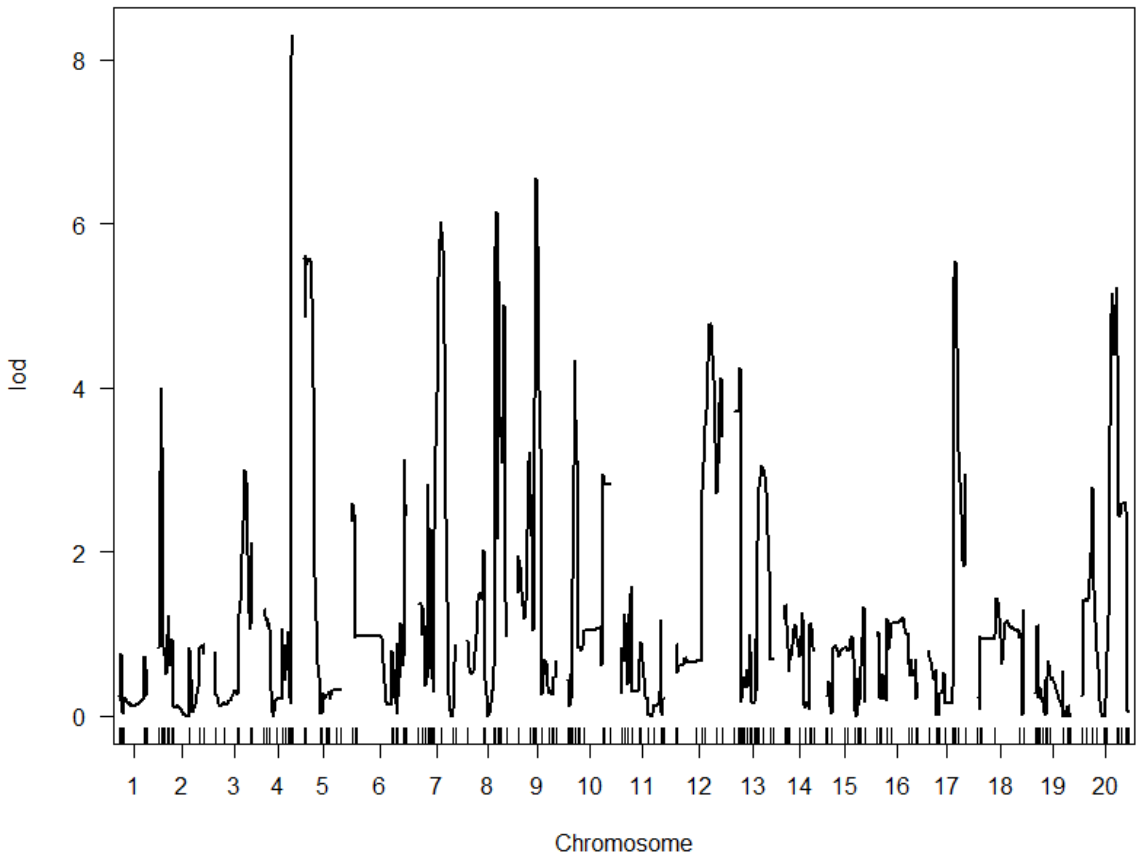


Figure 4. Manhattan plot showing the logarithm of odds (LOD) across soybean genome positions from the interval mapping QTL analysis using the highly significant QTL on Gm20 as a covariate in the analysis for protein content of BC_4F_4 derived lines.

.

Table 14. QTL with significance at the P<0.01 threshold for seed protein content including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
c2.loc4	2	22.4	11.60	0.0000
c3.loc64	3	76.5	8.00	0.0000
BARC-042189-08197	4	68.8	7.73	0.0000
c5.loc16	5	20.1	5.26	0.0003
BARC-019987-03748	7	64.7	6.41	0.0001
c8.loc78	8	108.7	6.17	0.0001
BARC-014813-01678	9	47.4	5.84	0.0002
c10.loc68	10	111.7	4.16	0.0053
c12.loc76	12	82.2	6.80	0.0001
BARC-028887-06033	13	76.7	9.38	0.0000
c14.loc22	14	25.3	6.82	0.0001
c16.loc22	16	24.3	3.99	0.0077
c17.loc58	17	79.3	6.79	0.0001
BARC-041129-07912	20	33.2	28.12	0.0000

Table 15. QTL with significance at the P<0.01 threshold for seed yield including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
c14.loc38	14	41.3	6.04	0.0010
c20.loc24	20	35.4	8.50	0.0000

Table 16. QTL with significance at the P<0.01 threshold for seed oil content including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
c2.loc4	2	22.4	13.85	0.0000
c3.loc64	3	76.5	8.10	0.0000
BARC-018283-03551	7	48.8	5.64	0.0009
BARC-014659-01609	9	34.0	5.01	0.0018
c10.loc66	10	109.7	5.57	0.0009
c12.loc98	12	104.2	5.71	0.0007
BARC-028887-06033	13	76.7	10.31	0.0000
c14.loc22	14	25.3	7.34	0.0001
c16.loc22	16	24.3	4.86	0.0022
BARC-019787-04375	17	78.5	4.57	0.0045
BARC-041129-07912	20	33.2	27.97	0.0000

Table 17. QTL with significance at the P<0.01 threshold for seed size (GPC) including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
BARC-064297-18613	6	129.5	8.62	0.0000
c8.loc78	8	108.7	5.47	0.0023
c9.loc28	9	37.3	6.87	0.0001
BARC-021603-04153	18	104.2	5.01	0.0053
BARC-051673-11193	19	94.0	5.49	0.0023

Table 18. QTL with significance at the P<0.01 threshold for maturity including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
c2.loc8	2	26.4	4.73	0.0023

Table 19. QTL with significance at the P<0.01 threshold for the sum of protein and oil content (P+O) including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
c2.loc4	2	22.4	6.58	0.0000
c3.loc64	3	76.5	5.83	0.0002
BARC-042189-08197	4	68.9	11.35	0.0000
c5.loc14	5	18.1	5.52	0.0006
BARC-064297-18613	6	129.5	7.18	0.0000
BARC-019987-03748	7	64.7	8.84	0.0000
c8.loc78	8	108.7	9.93	0.0000
BARC-014813-01678	9	47.4	10.56	0.0000
c10.loc82	10	125.7	5.36	0.0011
c12.loc74	12	80.2	8.20	0.0000
c13.loc14	13	35.0	8.38	0.0000
c14.loc2	14	5.3	4.80	0.0033
c17.loc60	17	81.3	9.19	0.0000
BARC-041129-07912	20	33.2	15.70	0.0000

Table 20. QTL with significance at the P<0.01 threshold for seed protein content using the BARC-041129-07912 marker at Gm 20, 33.2 cM as a covariate in the analysis including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
BARC-042189-08197	4	68.9	8.30	0.0002
BARC-019987-03748	7	64.7	6.03	0.0121
c8.loc62	8	92.7	6.15	0.0095
BARC-014813-01678	9	47.4	6.56	0.0046

Table 21. QTL with significance at the P<0.01 threshold for seed oil content after using the BARC-041129-07912 marker at Gm 20, 33.2 cM as a covariate in the analysis including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
c2.loc4	2	22.4	5.94	0.0017
BARC-029531-06209	10	58.5	4.93	0.0108
c12.loc98	12	104.2	4.81	0.0145

Table 22. QTL with significance at the P<0.01 threshold for seed size (GPC) after using the BARC-041129-07912 marker at Gm 20, 33.2 cM as a covariate in the analysis including chromosome, genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
BARC-064297-18613	6	129.5	9.70	0.0000
c8.loc78	8	108.7	6.43	0.0003
c9.loc28	9	37.3	7.27	0.0001
BARC-021603-04153	18	104.2	5.54	0.0023
BARC-051673-11193	19	94.0	5.38	0.0031
c20.loc68	20	79.4	4.98	0.0063

Table 23. QTL with significance at the P<0.01 threshold for the sum of protein and oil content (P+O) after using the BARC-041129-07912 marker at Gm 20, 33.2 cM as a covariate in the analysis including chromosome (Gm), genetic position (Pos.; cM), logarithm of odds (LOD), and the P-value based on permutation testing.

Marker/Interval Name	Chromosome	Pos. (cM)	LOD	P-value
BARC-042189-08197	4	68.8	11.00	0.0000
c5.loc8	5	12.1	4.98	0.0120
BARC-064297-18613	6	129.5	6.92	0.0002
c7.loc48	7	64.9	8.12	0.0000
c8.loc78	8	108.7	8.58	0.0000
BARC-014813-01678	9	47.4	11.56	0.0000
c12.loc70	12	76.2	6.47	0.0005
c13.loc14	13	35.0	8.09	0.0000
c17.loc60	17	81.3	7.36	0.0000
BARC-044361-08677	20	85.4	7.41	0.0000

Table 24. Putative seed protein content QTL effects across all MN environments tested. QTL peaks were SNP marker loci in all cases except for the chromosome 8 region; the marker nearest the peak, BARC-019299-03876, was used for analysis. Information included for each location by genome position combination includes the allele coded in terms of parental origin; A allele indicates a SNP inherited from the recurrent parent 'Evans', B allele indicates a SNP inherited from PI15296, and H denotes a heterozygote. The number of individuals (N) in the BC4F4 derived population with the allele is included as well as the mean protein value (%) of lines that possess the respective alleles, the standard error associated with the mean as well as the 95% confidence intervals (C.I.), the R^2 value of the marker association and the significance (P>F)of the marker association.

Environment	Chromosome	20			4			7			8			9		
	Pos. (cM)	33.2			68.8			64.7			94.2			47.4		
	Marker	BARC-041129-07912			BARC-042189-08197			BARC-019987-03748			BARC-019299-03876			BARC-014813-01678		
	Allele	A	H	B	A	H	B	A	H	B	A	H	B	A	H	B
Danv-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	40.4	41.5	43.8	41.9	41.7	39.4	41.7	42.2	39.7	41.9	41.1	39.7	41.9	42.1	39.5
	SE	0.1	0.2	0.2	0.2	0.5	0.4	0.2	0.4	0.4	0.2	0.5	0.4	0.2	0.3	0.3
	Lower 95% C.I.	40.1	41.1	43.4	41.5	40.7	38.7	41.4	41.4	38.8	41.5	40.0	38.9	41.6	41.5	38.9
	Upper 95% C.I.	40.7	41.9	44.2	42.2	42.6	40.2	42.0	43.0	40.5	42.2	42.1	40.5	42.3	42.7	40.1
	R^2		*0.60*			*0.21*			*0.14*			*0.15*			*0.26*	
	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
Danv-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	38.7	39.9	42.1	40.2	40.0	37.8	40.1	40.5	37.8	40.2	39.5	38.1	40.2	40.4	38.0
	SE	0.1	0.2	0.2	0.2	0.5	0.4	0.2	0.4	0.4	0.2	0.5	0.4	0.2	0.3	0.3
	Lower 95% C.I.	38.4	39.5	41.7	39.9	39.1	37.1	39.7	39.8	36.9	39.8	38.4	37.3	39.9	39.8	37.3
	Upper 95% C.I.	39.0	40.3	42.5	40.5	40.9	38.6	40.4	41.3	38.6	40.5	40.5	38.9	40.5	41.0	38.6
	R^2		*0.6*			*0.21*			*0.14*			*0.15*			*0.26*	
	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
Lamb-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	40.8	41.9	44.1	42.2	42.3	39.9	42.0	42.5	40.1	42.2	41.5	40.1	42.2	42.4	40.1

	SE	0.1	0.2	0.2	0.2	0.5	0.4	0.2	0.4	0.4	0.2	0.5	0.4	0.2	0.3	0.3
	Lower 95% C.I.	40.5	41.4	43.7	41.8	41.3	39.1	41.7	41.7	39.3	41.8	40.5	39.3	41.9	41.8	39.4
	Upper 95% C.I.	41.1	42.3	44.5	42.5	43.2	40.6	42.4	43.3	41.0	42.5	42.6	41.0	42.5	43.0	40.7
	R^2		*0.58*			*0.20*			*0.13*			*0.14*			*0.22*	
	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
Morr-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	39.4	40.5	43.0	41.0	40.7	38.1	40.8	41.3	38.1	41.0	39.9	38.3	41.0	41.2	38.4
	SE	0.2	0.3	0.2	0.2	0.5	0.4	0.2	0.4	0.5	0.2	0.6	0.5	0.2	0.4	0.4
	Lower 95% C.I.	39.0	40.0	42.6	40.6	39.6	37.3	40.5	40.4	37.2	40.6	38.8	37.4	40.6	40.4	37.7
	Upper 95% C.I.	39.7	41.0	43.5	41.3	41.7	38.9	41.2	42.2	39.0	41.3	41.1	39.2	41.4	41.9	39.1
	R^2		*0.54*			*0.23*			*0.20*			*0.18*			*0.24*	
	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
Morr-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	38.6	39.6	42.0	40.0	39.6	38.0	39.9	40.3	37.9	40.0	39.2	38.2	40.1	40.1	38.0
	SE	0.1	0.2	0.2	0.2	0.5	0.4	0.2	0.4	0.4	0.2	0.5	0.4	0.2	0.3	0.3
	Lower 95% C.I.	38.4	39.2	41.6	39.7	38.7	37.3	39.6	39.6	37.1	39.7	38.2	37.4	39.8	39.5	37.3
	Upper 95% C.I.	38.9	40.0	42.4	40.3	40.6	38.7	40.2	41.1	38.7	40.3	40.3	39.0	40.4	40.7	38.6
	R^2		*0.61*			*0.16*			*0.14*			*0.11*			*0.21*	
	P>F		*<.0001*			*<.0001*			*<.0001*			0.0003			*<.0001*	
Rose-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	41.3	42.4	44.4	42.6	42.7	40.7	42.5	43.0	40.8	42.6	42.6	40.9	42.6	42.9	40.8
	SE	0.1	0.2	0.2	0.2	0.4	0.4	0.2	0.4	0.4	0.2	0.5	0.4	0.2	0.3	0.3
	Lower 95% C.I.	41.1	42.0	44.0	42.3	41.8	40.0	42.2	42.2	40.0	42.3	41.6	40.1	42.3	42.3	40.2
	Upper 95% C.I.	41.6	42.8	44.8	42.9	43.6	41.4	42.8	43.7	41.6	42.9	43.6	41.7	43.0	43.5	41.5
	R^2		*0.57*			*0.17*			*0.12*			*0.11*			*0.18*	
	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
Rose-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	39.8	40.6	42.5	40.9	40.9	39.0	40.8	41.4	39.0	40.9	40.3	39.2	40.9	41.1	39.1
	SE	0.1	0.2	0.2	0.1	0.4	0.3	0.1	0.3	0.3	0.1	0.4	0.3	0.1	0.3	0.3
	Lower 95% C.I.	39.5	40.3	42.1	40.6	40.1	38.4	40.5	40.7	38.3	40.6	39.4	38.5	40.7	40.6	38.6
	Upper 95% C.I.	40.0	41.0	42.8	41.2	41.6	39.7	41.1	42.0	39.7	41.2	41.2	39.9	41.2	41.7	39.6
	R^2		*0.55*			*0.18*			*0.18*			*0.14*			*0.23*	

	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
Overall	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	39.9	40.9	43.1	41.2	41.1	39.0	41.1	41.6	39.0	41.2	40.6	39.2	41.3	41.5	39.1
	SE	0.1	0.2	0.2	0.2	0.4	0.4	0.2	0.4	0.4	0.2	0.5	0.4	0.2	0.3	0.3
	Lower 95% C.I.	39.6	40.5	42.8	40.9	40.2	38.3	40.8	40.9	38.2	40.9	39.6	38.4	41.0	40.9	38.5
	Upper 95% C.I.	40.1	41.3	43.5	41.5	42.0	39.7	41.4	42.3	39.8	41.5	41.6	40.0	41.6	42.1	39.7
	R^2		*0.61*			*0.21*			*0.17*			*0.15*			*0.24*	
	P>F		*<.0001*			*<.0001*			*<.0001*			*<.0001*			*<.0001*	
2008-PYT	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	39.8	41.1	43.2	41.3	41.2	39.2	41.1	41.9	39.2	41.2	40.4	39.6	41.3	41.5	39.2
	SE	0.2	0.3	0.3	0.2	0.6	0.5	0.2	0.5	0.5	0.2	0.7	0.5	0.2	0.4	0.4
	Lower 95% C.I.	39.4	40.4	42.6	40.8	39.9	38.2	40.7	40.9	38.1	40.8	39.0	38.5	40.8	40.6	38.3
	Upper 95% C.I.	40.2	41.7	43.8	41.7	42.4	40.2	41.5	42.9	40.2	41.7	41.8	40.7	41.8	42.3	40.1
	R^2		*0.39*			*0.09*			*0.10*			*0.06*			*0.13*	
	P>F		*<.0001*			*0.0015*			*0.0010*			*0.0144*			*0.0001*	

Table 25. Difference in soybean seed protein content (%) for lines with contrasting SNP alleles at putative protein content QTL across MN environments (allelic substitution). Values represent the difference in protein content between 'Evans' inherited allele (A) – PI153296 inherited allele (B). All differences are significant (P<0.05; Table 24).

Chromosome	20	4	7	8	9
Position (cM)	33.2	68.8	64.7	94.2	47.4
Marker	BARC-041129-07912	BARC-042189-08197	BARC-019987-03748	BARC-019299-03876	BARC-014813-01678
Environment	----------A-B----------				
Danv-2010	-3.44	2.45	2.00	2.17	2.43
Danv-2011	-3.37	2.35	2.29	2.09	2.23
Lamb-2010	-3.31	2.29	1.89	2.02	2.13
Morr-2010	-3.67	2.85	2.74	2.64	2.58
Morr-2011	-3.33	2.02	1.98	1.76	2.12
Rose-2010	-3.09	1.95	1.67	1.68	1.80
Rose-2011	-2.69	1.86	1.79	1.71	1.83
Overall	-3.27	2.26	2.08	2.03	2.17
2008-PYT	-3.43	2.03	1.92	1.64	2.10

Table 26. Putative seed protein content QTL yield effects across all MN environments tested. QTL peaks were SNP marker loci in all cases except for the chromosome 8 region; the marker nearest the peak, BARC-019299-03876, was used for analysis. Information included for each location by genome position combination includes the allele coded in terms of parental origin; A allele indicates a SNP inherited from the recurrent parent 'Evans', B allele indicates a SNP inherited from PI15296, and H denotes a heterozygote. The number of individuals (N) in the BC_4F_4 derived population with the allele is included as well as the mean yield value (Mg ha^{-1}) of lines that possess the respective alleles, the standard error associated with the mean as well as the 95% confidence intervals (C.I.), the R^2 value of the marker association and the significance (P>F) of the marker association.

Environment	Chromosome	20			4			7			8			9		
	Pos. (cM)	33.2			68.8			64.7			94.21			47.384		
	Marker	BARC-041129-07912			BARC-042189-08197			BARC-019987-03748			BARC-019299-03876			BARC-014813-01678		
	Allele	A	H	B	A	H	B	A	H	B	A	H	B	A	H	B
Danv-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.20	2.18	2.05	2.13	2.02	2.36	2.14	2.21	2.21	2.14	2.19	2.27	2.15	2.11	2.24
	SE	0.03	0.05	0.05	0.03	0.08	0.06	0.03	0.07	0.07	0.03	0.09	0.07	0.03	0.05	0.06
	Lower 95% C.I.	2.14	2.08	1.96	2.08	1.87	2.24	2.09	2.08	2.07	2.08	2.02	2.14	2.09	2.00	2.13
	Upper 95% C.I.	2.27	2.27	2.15	2.18	2.18	2.49	2.20	2.34	2.35	2.19	2.37	2.41	2.21	2.22	2.36
	R2		*0.05*			*0.10*			*0.01*			*0.03*			*0.02*	
	P>F		*0.0349*			*0.0008*			*0.4811*			*0.1750*			*0.2003*	
Danv-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.82	2.70	2.59	2.73	2.72	2.73	2.71	2.69	2.89	2.72	2.82	2.77	2.73	2.69	2.78
	SE	0.03	0.05	0.04	0.03	0.08	0.06	0.03	0.06	0.07	0.03	0.09	0.07	0.03	0.05	0.06
	Lower 95% C.I.	2.76	2.61	2.50	2.68	2.57	2.61	2.66	2.57	2.76	2.67	2.66	2.64	2.67	2.59	2.67
	Upper 95% C.I.	2.88	2.79	2.67	2.79	2.88	2.86	2.77	2.81	3.02	2.77	2.99	2.90	2.79	2.80	2.89
	R2		*0.14*			*0.00*			*0.05*			*0.01*			*0.01*	
	P>F		*<.0001*			*0.9871*			*0.0392*			*0.3970*			*0.5038*	

Lamb-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.80	2.78	2.67	2.76	2.85	2.77	2.74	2.83	2.87	2.76	2.81	2.76	2.75	2.80	2.79
	SE	0.04	0.06	0.05	0.03	0.09	0.07	0.03	0.07	0.08	0.03	0.10	0.08	0.03	0.06	0.06
	Lower 95% C.I.	2.73	2.67	2.56	2.70	2.67	2.63	2.68	2.69	2.71	2.70	2.61	2.61	2.68	2.68	2.66
	Upper 95% C.I.	2.88	2.89	2.77	2.82	3.03	2.91	2.80	2.97	3.02	2.82	3.00	2.91	2.81	2.92	2.91
	R2		*0.04*			*0.01*			*0.02*			*0.00*			*0.01*	
	P>F		*0.0895*			*0.6160*			*0.1941*			*0.8910*			*0.6524*	
Lamb-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.32	2.44	2.34	2.36	2.49	2.21	2.31	2.47	2.51	2.37	2.39	2.23	2.40	2.41	2.14
	SE	0.06	0.09	0.09	0.05	0.14	0.12	0.05	0.12	0.13	0.05	0.16	0.12	0.05	0.10	0.10
	Lower 95% C.I.	2.20	2.26	2.17	2.26	2.20	1.98	2.21	2.23	2.26	2.27	2.07	1.99	2.29	2.22	1.94
	Upper 95% C.I.	2.44	2.62	2.51	2.45	2.77	2.44	2.41	2.70	2.76	2.47	2.70	2.47	2.50	2.60	2.34
	R2		*0.01*			*0.02*			*0.02*			*0.01*			*0.04*	
	P>F		*0.5375*			*0.2967*			*0.2031*			*0.5470*			*0.0711*	
Morr-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	1.99	1.96	1.86	1.95	1.93	1.95	1.93	1.94	2.06	1.93	2.00	2.03	1.95	1.90	2.00
	SE	0.03	0.04	0.04	0.02	0.07	0.06	0.02	0.06	0.06	0.02	0.08	0.06	0.03	0.05	0.05
	Lower 95% C.I.	1.93	1.88	1.78	1.90	1.80	1.84	1.89	1.83	1.94	1.89	1.85	1.92	1.90	1.80	1.90
	Upper 95% C.I.	2.04	2.05	1.94	2.00	2.07	2.05	1.98	2.05	2.18	1.98	2.15	2.15	2.00	1.99	2.10
	R2		*0.05*			*0.00*			*0.03*			*0.02*			*0.02*	
	P>F		*0.0350*			*0.9740*			*0.1352*			*0.2188*			*0.3022*	
Morr-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.77	2.65	2.30	2.57	2.71	2.85	2.60	2.60	2.84	2.59	2.64	2.83	2.60	2.58	2.75
	SE	0.04	0.06	0.06	0.04	0.11	0.08	0.04	0.09	0.09	0.04	0.12	0.09	0.04	0.07	0.08
	Lower 95% C.I.	2.70	2.54	2.19	2.50	2.50	2.68	2.52	2.42	2.66	2.52	2.41	2.65	2.52	2.44	2.60
	Upper 95% C.I.	2.85	2.77	2.41	2.65	2.92	3.01	2.67	2.77	3.03	2.66	2.88	3.01	2.68	2.73	2.91
	R2		*0.27*			*0.07*			*0.04*			*0.04*			*0.02*	

	P>F		*<.0001*			*0.0108*			*0.0502*			*0.0481*			*0.1869*	
Rose-2010	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.26	2.24	2.15	2.21	2.29	2.28	2.21	2.25	2.34	2.21	2.33	2.28	2.20	2.21	2.32
	SE	0.02	0.04	0.03	0.02	0.06	0.05	0.02	0.05	0.05	0.02	0.06	0.05	0.02	0.04	0.04
	Lower 95% C.I.	2.21	2.17	2.08	2.17	2.18	2.19	2.17	2.15	2.24	2.17	2.20	2.19	2.16	2.14	2.25
	Upper 95% C.I.	2.31	2.31	2.21	2.25	2.41	2.37	2.24	2.34	2.44	2.25	2.45	2.38	2.25	2.29	2.40
	R2		*0.06*			*0.02*			*0.05*			*0.04*			*0.05*	
	P>F		*0.0199*			*0.1904*			*0.0418*			*0.0894*			*0.0291*	
Rose-2011	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.20	2.13	1.95	2.12	2.06	2.16	2.12	2.10	2.17	2.12	2.12	2.15	2.14	2.08	2.11
	SE	0.03	0.04	0.04	0.02	0.07	0.06	0.02	0.06	0.06	0.02	0.08	0.06	0.03	0.05	0.05
	Lower 95% C.I.	2.15	2.05	1.87	2.07	1.92	2.05	2.07	1.98	2.04	2.07	1.96	2.03	2.08	1.98	2.01
	Upper 95% C.I.	2.26	2.22	2.02	2.17	2.20	2.28	2.17	2.21	2.29	2.17	2.27	2.27	2.19	2.18	2.21
	R2		*0.18*			*0.01*			*0.01*			*0.00*			*0.01*	
	P>F		*<.0001*			*0.5075*			*0.6905*			*0.8970*			*0.6135*	
Overall	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	2.42	2.39	2.24	2.35	2.39	2.41	2.34	2.40	2.49	2.36	2.41	2.41	2.37	2.36	2.38
	SE	0.02	0.03	0.03	0.02	0.05	0.04	0.02	0.04	0.04	0.02	0.06	0.04	0.02	0.03	0.04
	Lower 95% C.I.	2.38	2.34	2.19	2.32	2.30	2.33	2.31	2.32	2.41	2.32	2.31	2.33	2.33	2.29	2.31
	Upper 95% C.I.	2.46	2.45	2.29	2.39	2.49	2.49	2.38	2.48	2.57	2.39	2.52	2.50	2.41	2.42	2.45
	R2		*0.18*			*0.02*			*0.08*			*0.02*			*0.00*	
	P>F		*<.0001*			*0.3443*			*0.0051*			*0.3266*			*0.8796*	
2008-PYT	N	71	31	35	105	12	19	103	18	16	110	10	17	87	26	24
	Mean	1.42	1.40	1.34	1.38	1.41	1.46	1.38	1.45	1.41	1.39	1.38	1.45	1.36	1.38	1.54
	SE	0.02	0.04	0.03	0.02	0.06	0.05	0.02	0.05	0.05	0.02	0.06	0.05	0.02	0.04	0.04
	Lower 95% C.I.	1.37	1.33	1.28	1.35	1.30	1.37	1.34	1.35	1.32	1.35	1.26	1.35	1.32	1.31	1.47
	Upper 95% C.I.	1.46	1.48	1.41	1.42	1.52	1.55	1.42	1.54	1.51	1.43	1.51	1.54	1.40	1.45	1.62

R2	*0.02*	*0.02*	*0.01*	*0.01*	*0.12*
P>F	*0.2033*	*0.3271*	*0.4184*	*0.4976*	*0.0002*

Table 27. Difference in soybean seed yield (Mg ha^{-1}) for lines with contrasting SNP alleles at putative protein content QTL across MN environments (allelic substitution). Values represent the difference in yield between 'Evans' inherited allele (A) – PI 153296 inherited allele (B). Differences are significant where denoted.

Chromosome	20		4		7		8		9	
Position (cM)	33.2		68.8		64.7		94.2		47.4	
Marker	BARC-041129-07912		BARC-042189-08197		BARC-019987-03748		BARC-019299-03876		BARC-014813-01678	
Environment	----------A-B----------									
Danv-2010	0.15	*	-0.23	***	-0.07	NS	-0.13	NS	-0.10	NS
Danv-2011	0.23	***	0.00	NS	-0.18	*	-0.05	NS	-0.05	NS
Lamb-2010	0.14	NS	-0.01	NS	-0.13	NS	0.00	NS	-0.04	NS
Lamb-2011	-0.02	NS	0.15	NS	-0.20	NS	0.14	NS	0.26	NS
Morr-2010	0.13	*	0.00	NS	-0.13	NS	-0.10	NS	-0.05	NS
Morr-2011	0.47	***	-0.27	*	-0.25	*	-0.24	*	-0.15	NS
Rose-2010	0.11	*	-0.07	NS	-0.13	*	-0.08	NS	-0.12	*
Rose-2011	0.26	***	-0.05	NS	-0.05	NS	-0.03	NS	0.02	NS
Overall	0.18	***	-0.06	NS	-0.15	**	-0.06	NS	-0.01	NS
2008-PYT	0.07	NS	-0.07	NS	-0.03	NS	-0.06	NS	-0.18	***

***Significant at P<0.001

*Significant at P<0.05

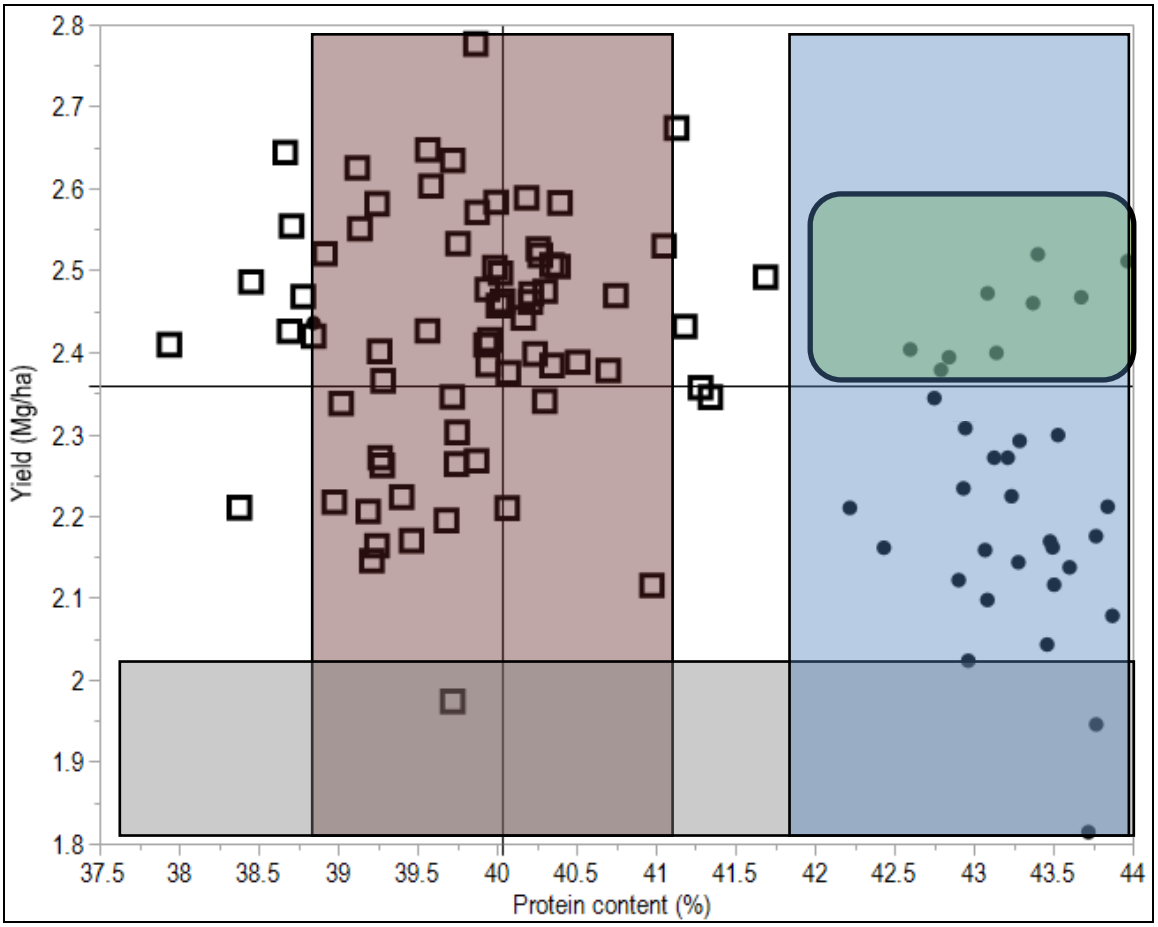


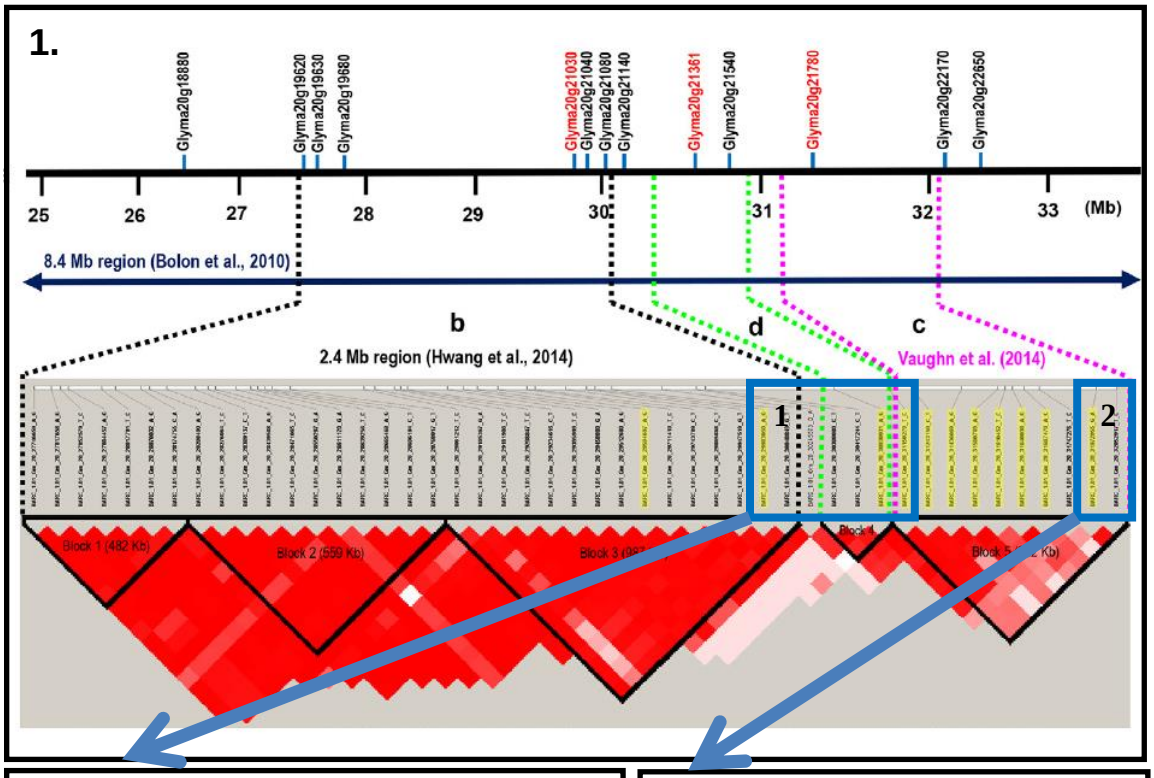
Figure 5. Combined least square means for seed yield (Mg ha^{-1}) vs seed protein concentration (% dry basis) for BC_4F_4 derived lines developed from the recurrent parent 'Evans' and the high protein donor parent 'PI153296'. The solid lines originating on the Y-axis (2.36) and X-axis (40.03) denote the mean value for yield and protein of the recurrent parent. The red shaded box indicates the 95% confidence interval around the mean protein concentration of 'Evans'. The blue shaded box denotes the lines that possess significantly (α=0.05) greater protein than 'Evans'. The gray shaded box indicates the lines that have significantly lower yield than 'Evans'. Lines indicated with open squares denote lines with the Evans allele and closed circles indicate lines with the PI296153at SNP marker BARC-044361-08677 at chromosome 20, 33.2 cM.

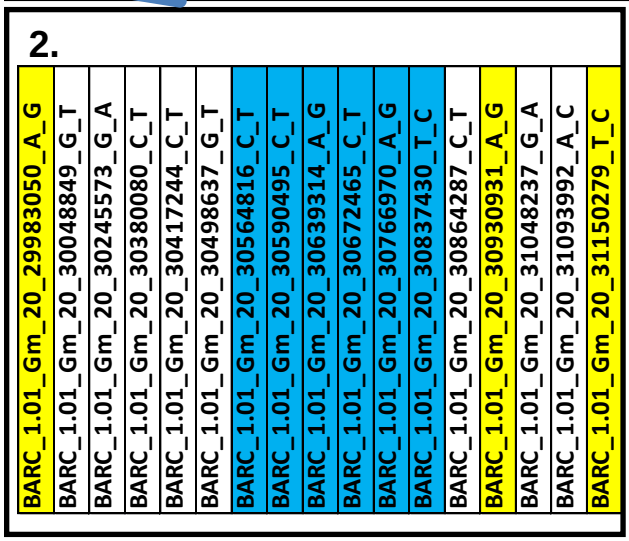
Table 28. Haplotype frequency of Evans and PI153296 for haplotype block regions 4 and 5 defined by Bandillo et al., 2015 and corresponding protein content from data obtained through the Soybean GRIN database.

			N		Mean	Lower 95% C.I.	Upper 95% C.I.
Haplotype block †	Position Mbp	Haplotype	Total	with Protein		Protein (%)‡	
4	30.38-30.93	Evans	18900	14757	44.21	44.17	44.25
		PI153296	120	29	49.52	48.56	50.47
5	31.15-32.05	Evans	9591	7540	43.80	43.75	43.87
		PI153296	687	416	45.72	45.50	45.95

†Bandillo et al., 2015

‡Mean protein differences for the two haplotypes are significant ($P<0.001$) different for both haplotype blocks


1.
Glyma20g18880
Glyma20g19620
Glyma20g19630
Glyma20g19680
Glyma20g21030
Glyma20g21040
Glyma20g21080
Glyma20g21140
Glyma20g21361
Glyma20g21540
Glyma20g21780
Glyma20g22170
Glyma20g22650
25
26
27
28
29
30
31
32
33
(Mb)
8.4 Mb region (Bolon et al., 2010)
b
d
c
2.4 Mb region (Hwang et al., 2014)
Vaughn et al. (2014)
1
2
Block 1 (482 Kb)
Block 2 (559 Kb)


2.
BARC_1.01_Gm_20_29983050_A_G
BARC_1.01_Gm_20_30048849_G_T
BARC_1.01_Gm_20_30245573_G_A
BARC_1.01_Gm_20_30380080_C_T
BARC_1.01_Gm_20_30417244_C_T
BARC_1.01_Gm_20_30498637_G_T
BARC_1.01_Gm_20_30564816_C_T
BARC_1.01_Gm_20_30590495_C_T
BARC_1.01_Gm_20_30639314_A_G
BARC_1.01_Gm_20_30672465_C_T
BARC_1.01_Gm_20_30766970_A_G
BARC_1.01_Gm_20_30837430_T_C
BARC_1.01_Gm_20_30864287_C_T
BARC_1.01_Gm_20_30930931_A_G
BARC_1.01_Gm_20_31048237_G_A
BARC_1.01_Gm_20_31093992_A_C
BARC_1.01_Gm_20_31150279_T_C

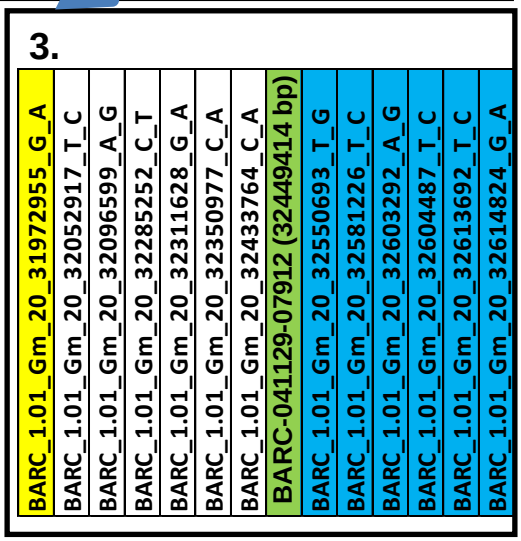

3.
BARC_1.01_Gm_20_31972955_G_A
BARC_1.01_Gm_20_32052917_T_C
BARC_1.01_Gm_20_32096599_A_G
BARC_1.01_Gm_20_32285252_C_T
BARC_1.01_Gm_20_32311628_G_A
BARC_1.01_Gm_20_32350977_C_A
BARC_1.01_Gm_20_32433764_C_A
BARC-041129-07912 (32449414 bp)
BARC_1.01_Gm_20_32550693_T_G
BARC_1.01_Gm_20_32581226_T_C
BARC_1.01_Gm_20_32603292_A_G
BARC_1.01_Gm_20_32604487_T_C
BARC_1.01_Gm_20_32613692_T_C
BARC_1.01_Gm_20_32614824_G_A

Figure 6. Pane 1. Chromosome 20 region defined in Bandillo et al. (2016). Markers shaded yellow in panes 2 and 3 correspond to the yellow markers indicated in pane 1. Blue shaded markers indicate markers that are polymorphic SoySNP50K markers between 'Evans' and PI153296. The green marker indicated in pane 3 is the position of the BARC-041129-07912 SNP marker on the GoldenGate assay which was found to be the most significant marker in the current study. All non-blue and green markers are monomorphic between the 'Evans' and PI153296. Pane 2 shows all of the SoySNP50K markers in the 29.98-31.15 Mbp region on chromosome 20. While pane 3 shows the SoySNP50K markers in the 31.97-32.61 Mpb region with the GoldenGate marker positioned in between. Haplotype block 4 defined by Bandillo et al. (2016) contains the segregating region in the current study (pane 2), while the segregating region shown in pane 3 is slightly downstream of haplotype block 5.

Chapter 3 - Soybean Yield Response to Plant Population Density among Full-Season and Short-Season Cultivars across Diverse Environments

Outline

Much variation exists in the literature for soybean yield response to seed/plant density. One likely contributing factor to this variation is the discontinuity of experiments targeted at addressing this question across the soybean agronomic research community (e.g. research is generally carried out on a state by state basis). Other factors such as the environments and geographies evaluated, the specific cultivars studied, and the cultural practices implemented in the experiments also contribute to this variability. A driving force behind this report was to bring a group of soybean production specialists across the United States together to address this fundamental question of optimum plant density with an experimental design that was consistent across the entire range of environments sampled. We hypothesized that soybean yield response to plant population will differ depending on cultivar maturity and latitude and that adapted, full-season cultivars would require a reduced plant density to achieve maximum yield relative to short-season cultivars. To test our hypotheses, we evaluated soybean yield response to seeded density and harvested plant density for full-season and short-season cultivars across 59 environments distributed across the United States from 30.2° to 47.8° N. latitude. We observed that full-season cultivars yielded greater than shorter-season cultivars at 23 of the 25 environments that showed a significant yield difference $(\alpha=0.05)$ for cultivar maturity. However, short-season cultivars can achieve comparable yields when seeded at higher densities compared to full-season cultivars. We observed more northerly locations to require greater plant densities to achieve 95% of asymptotic yield (Y95%) compared to southern environments.

Introduction

Since the inception of wide spread soybean (*Glycine max* (L. Merr) cultivation in the United States in the mid 1900's, crop scientists have worked to evaluate cultural practices that maximize seed yield and economic return. The implementation of crop cultural practices is among the few factors associated with crop productivity that can be controlled by the producer; thus, production practice decisions are the crux of farmer profitability. As such, crop management decisions should be based on up-to-date, reliable agronomic research. Numerous experiments addressing important production issues such as optimum planting date (Anderson and Vasilas, 1985; Beuerlein, 1988; DeBruin and Pedersen, 2008a; Egli and Cornelius, 2009; Elmore, 1990; Grau et al., 1994; Lee et al., 2008; Lueschen et al., 1992; Oplinger and Philbrook, 1992; Pedersen and Lauer, 2004; Wilcox and Frankenberger, 1987), row spacing (Ablett et al., 1991; Alessi and Power, 1982; Beuerlein, 1988; Bullock et al., 1998; Cooper, 1977; DeBruin and Pedersen, 2008b; Devlin et al., 1995; Egli, 1988; Egli, 1994; Elmore, 1991; Ethredge et al., 1989; Grau et al., 1994; Holshouser and Whittaker, 2002; Janovicek et al., 2006; Kratochvil et al., 2004; Lee, 2006; Lueschen et al., 1992; Oplinger and Philbrook, 1992; Pedersen and Lauer, 2003; Taylor, 1980; Weber et al., 1966; Wiggans, 1939), and plant population density (Ablett et al., 1991; Beuerlein, 1988; Carpenter and Board, 1997; Costa et al., 1980; Cox and Cherney, 2011; DeBruin and Pedersen, 2008a,b; Devlin et al., 1995; Edwards and Purcell, 2005; Egli, 1988; Elmore, 1991; Elmore, 1998; Ethridge et al., 1989; Lee et al., 2008; Leffel and Barber, 1961; Lehman and Lambert, 1960; Lueschen and Hicks, 1977; Parks et al., 1982; Kratochvil et al., 2004; Oplinger and Philbrook,

1992; Pedersen and Lauer, 2002; Walker et al., 2010; Weber et al., 1966; Wells, 1991) have been reported by numerous crop scientists.

As it relates to the present work, seed density at planting is a key decision made annually by producers. Because population density is one of the fundamental production decisions, soybean growth and yield response to varying seeding densities and plant populations has been a well-studied topic within the agronomic sciences community. Since the first Agronomy Journal report in 1939 (Wiggens) on the topic, many researchers have assessed the seeding rates and plant populations required to achieve maximum yield (*see previous citation list for plant population density studies*). These reports describe mixed results, and as a consequence, the recommended seeding rate or final plant stand that results in optimum yield is varied. Generally, soybean yield increases curvilinearly as plant population increases until a maximum or plateau is reached. Maximum yield has been observed with final plant stands from 7 plants m^{-2} to 49 plants m^{-2} (Costa et al., 1980; Egli, 1988; Leffel and Barber, 1961; Lehman and Lambert, 1960; Lueschen and Hicks, 1977; Wells, 1991). However, it has been observed that for a single variety grown at one location, optimum plant population can vary by 100% or more between years (Moore and Longer, 1987; Wells, 1991). Such variation can be attributed to the compensatory nature of the soybean plant, and yield maintenance across varying plant densities is due primarily to the modulation of branch yield components (Carpenter and Board, 1997; Lueschen and Hicks, 1977). The variability in yield response to population can also be attributed to growing conditions during a given

season; with the minimal optimal plant population being lower under more favorable growing conditions (Devlin et al., 1995; Elmore, 1998; Wells, 1993).

In addition to environment growing conditions, latitude appears to play a role in the optimum recommended plant density. Southern latitudes tend to require fewer plants per area compared to northerly locations (Lee et al., 2008; Heatherly and Elmore, 2003; DeBruin and Pedersen, 2008). Differences between northern and southerly environments have been attributed to the differences in accrual of thermal units and photosythetically active solar radiation (PAR) when considering adapted cultivars to the respective region (Lee, 2006; Edwards et al., 2005). Cultivar maturity plays a similar role to latitude in influencing yield response to plant population by altering the period for which the crop is accumulating thermal units and PAR (Edwards and Purcell, 2005). Further differences in results can be attributed to the era when the experiment was conducted; modern cultivars have been observed to be more responsive to increased seeding rates than older cultivars (Suhre et al., 2014). Lastly, inherent genetic differences between the varieties studied, row width, cropping system, and site-years of data could explain the variability in reported optimum soybean seeding rates.

Collectively, the aforementioned factors are likely the basis for the differences in seeding rate recommendations by region and state. To exemplify the differences in recommended seeding rates; current Minnesota seeding rate recommendations range from 34.6 – 42.0 seeds m^{-2} (Naeve, 2008) depending on cultivar maturity and environmental factors; Iowa seeding rate recommendations range from 30.9 – 34.6 seeds m^{-2} (DeBruin and Pedersen, 2007); Michigan seeding rate recommendations range from

32.1 – 43.2 seeds ha^{-1} (Stanton et al., 2011); Missouri recommendations range from 32.1 – 49.4 seeds m^{-2} (Helsel and Minor, 1993); Illinois recommends 37 seeds m^{-2} (Davis, 2010); Arkansas recommends 44.5 seeds m^{-2} for maturity group III and IV indeterminate cultivars (Ashlock et al., 2007).

Although state to state variation in seeding rates exist, it is generally accepted that 24.7 plants m^{-2} at harvest should be sufficient to obtain maximum yield for an environment (DeBruin and Pedersen, 2008; Lee et al., 2008; Naeve, 2008; DeBruin and Pedersen, 2007; Lee and Herbeck, 2011; Davis, 2010). Ultimately, seeding rate recommendations are designed to establish a minimum stand that results in optimum yield. Differences in stand mortality, or seed/plant loss from emergence until harvest understandably impacts the seeding rate required to achieve a target final plant stand. Seeding rate recommendations are based on pure live seeds; the quotient of target seeding density and seed lot germination. Even with germination correction, 100% of the pure live seeds generally do not emerge. Several factors including compaction (Hyatt et al., 2007), soil type (Yaklich et al., 1979), tillage system (Vetch et al., 2007), soil-borne pathogens (Hamman et al., 2002), soil temperature (Hatfield and Egli, 1974; Unander et al., 1986; Helms et al., 1996), moisture status (Helms et al., 1996), and seeding depth (Fehr et al., 1973) have been shown to influence field emergence of soybean. Seedling losses from planting to emergence have been reported in the range of 5% - 30% (Cox and Cherney, 2011; Kratochvil et al., 2004; Norsworthy and Fredrich, 2002). In addition to losses from seeding to emergence, losses from emergence to harvest have been reported in the range of 2.5% - 38% (Cox and Cherney, 2011; Board, 2000). Cumulative stand

losses from planting reported in the literature range from 7% (Cox and Cherney, 2011) to 20-40% (DeBruin and Pedersen, 2008b; Ethridge et al., 1989; Lee et al., 2008; Walker et al., 2010;). Greater stand mortality is typically observed in wide-row widths (i.e. >76 cm) compared to narrower row spacings which is attributed to greater interplant competition (Oplinger and Philbrook, 1992; DeBruin and Pedersen, 2008b).

In an attempt to work toward streamlining United States soybean plant density recommendations, we hypothesized that soybean yield response to plant population will differ depending on cultivar maturity and latitude; in that northerly U.S. locations will be more responsive to plant population than southerly U.S. locations, and adapted, full-season cultivars would require a reduced plant density to achieve maximum yield relative to short-season cultivars. To test our hypotheses, the objectives of this study are to evaluate soybean yield response to seeded density and harvested plant density across multiple environments distributed across the United States from 30.2° to 47.8° N. latitude. Stand attrition from the seeded, emerged, and harvested density will be assessed.

Materials and Methods

Field experimentation

Experiments were conducted at a total of 59 site-year combinations during the 2009, 2010, and 2011 growing seasons in Minnesota, Michigan, Iowa, Kentucky, Arkansas, and Louisiana. Information regarding experimental locations is included in Table 1. The experiments in each environment were arranged as a randomized completed block design in a split-plot arrangement with four blocks (replications). Main plots were cultivar maturity (2): a full-season cultivar and a short-season cultivar approximately one maturity group earlier than adapted (Table 1). The split plot was seeding density (6): 6.2, 18.5, 30.9, 43.2, 55.6, and 67.9 seeds m^{-2}. Plots were at least 1.5 m wide by 6 m long and were planted in narrow row spacing (51 cm or less) at all locations except Louisiana where plots were planted on 96 cm spaced raised beds with a single row per bed to facilitate drainage and irrigation. Soil was sampled at each environment and fertilized to the recommendations for high-yields (>4.5 Mg ha^{-1}) according to each cooperating university prior to each season. Weeds were controlled with glyphosate [N-(phosphonomethyl)glycine] and through manual weeding as necessary. The target planting date for each environment was to occur within the location's optimum planting window. Irrigation to supplement precipitation was supplied when accessible at experimental sites and irrigation timing and amount was applied based on each cooperating state's university recommended method.

Early season, established plant stand data were collected by counting all plants within a minimum of 1.5 m^2 area from each plot at the V1 stage of growth (Fehr and

Caviness, 1977). Harvested plant population data were collected similarly to emerged plant stands at the R8 stage of growth (Fehr and Caviness, 1977). Grain yield, corrected to 130 g kg^{-1} H_2O, was measured from a bordered area of each plot harvested with a plot combine. Minimum harvested area within each plot for yield was 9 m^2 at each site-year.

Data analysis

Soybean yield data were subjected to mixed model analysis using JMP software (SAS Institute, Cary, NC) where cultivar maturity designation, seeded density, and their interaction were considered fixed effects. Random effects included environment (each site-year combination), replication nested within environment, and all interactions therein. The amount of total variation associated with the random effects was determined using covariance parameter estimates. The purpose of the mixed model analysis was to draw overall conclusions on the potential effect and interaction of soybean maturity group with seeding density as well as to assess the amount of random variation that can be attributed to environments and interactions with environment. Soybean yield (Y) response to harvested plant density was modeled with a modified Mitscherlich equation with the NLIN procedure in SAS (Shabenberger and Pierce, 2003; SAS Institute Inc., Cary, NC, USA):

$$Y = \alpha\ (1\text{-}e^{-\beta x}) \qquad [1]$$

where α denotes the asymptote and β represents the responsiveness of grain yield to plant population. This model has been used in the past to explain soybean yield response to various crop inputs with success (Edwards et al., 2005; Board and Modeli, 2005; Lee et al., 2008; DeBruin and Pedersen, 2008). This model constrains the intercept to 0 for

practical purposes, as a plant population of 0 results in no seed yield. In some soybean yield by population studies, yield decreases have been observed at ultra-high populations often attributed to plant lodging (Cooper, 1971) making the appropriate model in those cases a quadratic or hyperbolic response; however, the greatest seeded density under investigation in the present study (67.9 seeds m^{-2}) were not high enough to observe lodging differences between the varying seeding rates of each cultivar (data not shown).

Separate models (Eq. 1) were fit for each site-year combination with separate α (asymptotes) and β (responsiveness) parameters for the two cultivars. Cultivar parameters were tested for differences by evaluating the 95% confidence intervals for each parameter. If confidence intervals overlapped for the two cultivars at a location, a model with a single α and/or β parameter was fit for that particular site-year. To confirm the confidence interval comparison method, the reduced model was compared to the full parameterized model using the sum of squares reduction test which tests for difference in the variation explained by the two models (Schabenburger and Pierce, 2003):

$$F_{obs} = \{(SSE_{Reduced} - SSE_{Full})/ (DFE_{Reduced} - DFE_{Full})\}/MSE_{Full} \quad [2]$$

Where SSE is the sum of squares error from the appropriate model, DFE is the degrees of freedom for the model error term, and MSE is mean square error. The test statistic from the sum of squares reduction test is compared to the threshold from an F-distribution which is determined by calculating the loss in degrees of freedom between the reduced and full models and dividing by the degrees of freedom in the full model. The P-value of the test described above was accomplished using the PROBF function in SAS.

If cultivar parameters were found to be different within site-year, then separate cultivar by year parameters were included for each. Once *cultivar* comparisons at each site-year were completed, a model describing yield for each *site* across years was constructed which included separate α and β parameters for each year and cultivar if found to be different based on confidence interval comparisons and sum of squares reduction testing as described previously. If β parameters across years for a given location appeared to be equivalent based on confidence interval comparisons, by site across year reduced model was compared to the full model using the sum of squares reduction test. If the models were found to be different, the more parameterized model was used. The models describing the relationship of seed yield to harvested plant population at each location over years goodness-of-fit was assessed with the pseudo-R^2 statistic calculated with the following equation.

$$Pseudo\text{-}R^2 = 1 - SS_{residual}/SS_{corrected\ total} \quad [2]$$

Pseudo-R^2 calculation is preferred to the standard R^2 statistic (coefficient of determination) for evaluating model fit in non-linear models and models that lack an intercept (Schabenberger and Pierce, 2003). For each model, optimum levels of harvested plant density were calculated for the density required to achieve 99% (Y99%) and 95% (Y95%) of the asymptotic yield maximum with the following equation:

$$PD = LN\ (1\text{-}\%goal)/\text{-}\beta \quad [3]$$

Where the natural log (LN) of one subtracted by the percentage of the asymptotic yield goal is divided by the responsiveness coefficient from Eq. 1. While Y99% yield was

calculated and reported, given the characteristics of the natural log function, multiplying the PD required for Y95% by 1.53 provides the PD required for Y99%.

To asses overall stand mortality across cultivars, seeded rates, and environments, a mixed effects model analysis of covariance was carried out. Cultivar, seed/plant density (on a continuous scale), and their interaction were considered fixed. Environment, replication, and any interaction with environment and replication were considered random effects. The model was fit to each of three pairwise combinations of stand density data: (i). seeded and emerged density, (ii). emerged and harvested density and, (iii). seeded and harvested density; all of which include cultivar maturity as the covariate. In addition to the all-inclusive model, a 'by environment' model was fit to determine the variability in stand establishment rates across the different environments sampled.

Results and Discussion

Yield response to seeded density

Environment had a large effect on soybean yield. The main effect of environment accounted for 64% of the total random variation. Replication nested within cultivar accounted for a total of 4% of the random variability. Cultivar and seeded rate interactions with environment accounted for 8.5% and 9% of the random variation, respectively. Fixed effects tests indicate that yield was significantly different across cultivar maturity designation, seeded density, and their interaction ($P<0.01$). Full-season cultivars, on average, produced a yield of 3.53 Mg ha^{-1} while short-season cultivars produced a yield of 3.29 Mg ha^{-1} ($LSD_{\alpha=0.05}=0.12$). Figure 1 displays the mean yield difference between full-season and short-season cultivars at each environment tested as determined via a 'by location' analysis. Numerically, later maturing cultivars outperformed earlier maturing cultivars at 70% of the environments (41 out of 59). At 23 of the 25 locations that indicated significant ($P<0.05$) differences for maturity main effects, the full-season cultivar significantly outpaced the short-season cultivar.

The lowest seeding rate at which yield was maximized in full-season cultivars occurred at a seeded rate of 43.2 seeds m^{-2}. This rate was not significantly different than the greater seeded densities of 55.6 and 67.9 seeds m^{-2} ($LSD_{\alpha=0.05}=0.124$). However, while soybean yield did not differ statistically for seeded rates above 43.2 seeds m^{-2}, there was a numerical increase in yield for the 55.6 and 67.9 seeds m^{-2} density. The 55.6 seeds m^{-2} density yielded 0.04 MG ha^{-1} greater than 43.6 seeds m^{-2}, and the 67.9 seeds m^{-2} density yielded 0.11 Mg ha^{-1} greater than 43.6 seeds m^{-2}.

Evaluation of the significant interaction between cultivar maturity designation and seeded density revealed that while full-season cultivars yielded the greatest, the short-season cultivars grown at high densities yielded similarly to the high and mid-range seeded densities for the full season cultivars as shown in Figure 2. Across all environments, short-season cultivars planted at 67.9 seeds m^{-2} did not differ significantly for yield compared to the full-season cultivar sowed at 67.9, 55.6, 43.2, and 30.9 seeds m^{-2}.

Yield response to harvested plant density and cultivar maturity

General overview

To evaluate the relationship between yield response to plant density across varying latitudes, plant density at which Y95% was reached for each site-year by cultivar combination was regressed against latitude. The cultivar and cultivar by plant population at 95% yield interaction was tested; however, neither term was significant (P=0.90 and 0.93, respectively). Figure 3 shows the results of the latitude x Y95% regression analysis which indicates a significant response (P=0.02). Plant populations required to achieve Y95% was greater at northerly latitudes than at more southerly locations. While the model fit as measured by the coefficient of determination was quite low (R^2=0.05), the significant model fit indicates that the overall trend of greater latitudes requiring a higher plant density compared to lower latitudes was observed. This trend displayed in Figure 3 appears to be driven, in part, by the high plant densities required in the northern MN environments and the lower plant densities required in the LA environments.

Location specific reduced models (Eq.1) fit the yield response to harvested plant density well with an average pseudo-R^2 of 0.74. Although model fits were generally good, similar to previous findings, yield response to final plant population was variable across environments and among cultivars grown within the same environment. Due to this variation, presenting the results in a simple concise way was not possible. The parameters α and β were initially fit for the complete set of 118 environment by cultivar combinations via Eq.1 (data not shown; although the data are represented in Figure 3). Parameter testing reduced the possible 118 unique models to 92 (Table 2). Location model goodness of fit for each location was generally good with pseudo-R^2 values typically greater than 0.7. Figures 10-12 display examples of environments affecting the parameter coefficients in Eq.1. Data presented in Figure 4 displays the yield response to plant density for two cultivars differing in maturity that were grown at Rosemount, MN during the 2010 growing season. The two cultivars have different yield levels, indicated by the different α parameters, but share the same responsiveness parameter coefficient (β=0.027) which indicates the optimum plant density required to achieve optimum yield was the same. This scenario occurred at 21 of the 59 site-years. Figure 5 displays the observed effect when yield levels are the same for two different cultivars, but differ in their responsiveness functions. In this example taken from Kalamazoo, MI from the 2009 growing season, the yield potential of both cultivars was 2.49 Mg ha^{-1}, but the plant population required to achieve optimal yields are different indicated by β parameters of 0.017 for the adapted cultivar and 0.005 for the earlier than adapted cultivar. This resulted in an optimum plant density to achieve Y95% at 66.7 and 17.2 plants m^{-2} for the

short-season and full-season cultivars, respectively. This scenario occurred nine out of 59 site-years. Results from Story City, IA are presented in Figure 6 and depict a single cultivar's variation for two growing seasons, 2009 and 2011. In this case, the same full-season cultivar had differing α's and β's. Yield was greater in 2009 than 2011, and the plant population required to achieve the greater yield in 2009 (19.1 plants m^{-2} at Y95%) was less than that required to obtain the lower yield realized in 2011 (37.7 plants m^{-2} at Y95%). This scenario of cultivars exhibiting differential responses for both yield and responsiveness occurred at three of the 59 site-years.

Of the 59 environments sampled, 34 showed no difference for yield potential between full-season and short season cultivars (Table 3). Of these 34; 24 were found to have no difference in the plant density responsiveness parameter β, and eight environments were identified as having differing β's. Short-season cultivars would be expected to have greater responsiveness to that of full-season cultivars; however, the distribution of these eight cases was 50:50. The adapted, full-season cultivar required a greater plant density than the short-season cultivar to obtain Y95% at Tuscola, MI in 2009 and 2010, Corning, IA in 2009, and Colt, AR in 2011. At two of the 59 environments the β parameter failed to converge, indicative of no difference in yield across the plant densities evaluated.

Twenty-five of the 59 environments showed significant yield differences between the two maturities for yield potential. The full-season cultivar had a greater α parameter at 16 environments compared to the short-season cultivar. Of the six environments where the short-season cultivar yielded greater than the full-season, four environments were

from Louisiana, where the designated short-season cultivar was indeterminate and the full-season cultivar was determinate. The determinate cultivar yielded greater than the indeterminate at only one Louisiana environment; Winnsboro in 2009. In all other Louisiana environments, the indeterminate cultivar out-yielded, or was similar to the determinant cultivar. The remaining two of eight environments where the short-season cultivar outperformed the full season cultivar were from St. Paul and Rosemount, MN in 2011. This was likely due to later than normal, planting dates at these two locations. The full season cultivar's yield at these two locations was affected negatively by a frost (e.g. ambient air temperature <0 C) September 14 (Coulter, 2011). Disregarding these anomalies, when differences in cultivar performance are observed, the full-season varieties had greater yield than short season cultivars.

While cultivar yield potential, α, is crucial information required by producers when making cultivar selection decisions, the plant population and ultimately seeding rate to implement is based on the β responsiveness coefficient (Eq. 1 and 3). Two cultivars that differ for α, yet share the same β, would respond the same to plant population (Figure 4). Only eleven of the 59 environments tested showed differences for β indicating that, at more than 80% of the environments there was no difference between full-season and short-season cultivars for yield response to plant population. Of the 11 instances that showed different β values, the short-season cultivar was found to be more responsive (smaller β value) to plant density in seven, which agrees with previous findings (Edwards and Purcell, 2005). The four environments where the full-season cultivar was more responsive to plant population were Tuscola, MI in 2009 and 2010,

Corning, IA in 2009, and Colt, AR in 2011. In each of these four cases, there was no difference in the asymptote parameter, thus no yield difference was observed between the two cultivars differing in maturity.

Comparisons among states

Given plant population recommendations have been, and will likely continue to be, delivered by soybean production agronomists on a by state basis; a results summary for each state where the current study was carried out is provided below.

Minnesota

We hypothesized that plant densities required to achieve optimum yields would differ based on a north-to-south gradient, with greater plant populations required for more northerly locations compared to southern locations. The northern most environments, Crookston and Morris, generally agree with our hypothesis in that these locations had the highest required plant densities to achieve optimal yields (Table 3). Average optimum plant density for these northern locations was 40.6 plants m^{-2} at the Y95% yield level and 62.3 plants m^{-2} at the Y99% yield level.

Central and southern Minnesota environments had an optimal plant density range between 7.5 – 20.4 plants m^{-2} (Y95% threshold) depending on environment and cultivar maturity. Although this range was rather wide, these environments achieved optimal yield between 11 – 13 plants m^{-2} (Y95%) for 19 out of the possible 24 environment by cultivar combinations (Table 3). Average optimum plant density for these central and southern locations was 12.5 plants m^{-2} at the Y95% level and 19.4 plants m^{-2} at the Y99% yield level.

Out of the 16 Minnesota environments, parameter tests for differences in individual cultivar response to plant density revealed only two instances where the optimum plant population was different for the two cultivars evaluated (Becker, 2010 and Waseca, 2009). In both these cases, the optimum plant density for the earlier maturing cultivar was greater than that required by the full season, adapted cultivar.

Michigan

Optimum plant densities in Michigan environments ranged from 8 plants m^{-2} at Tuscola in 2011 to 66.7 plants m^{-2} at Kalamazoo for the short season cultivar in 2009 for the Y95% threshold (Table 2). Our hypothesis predicted that Tuscola, the northerly most MI environment would require a greater plant density than East Lansing or Kalamazoo, yet this was not the case. On average, Tuscola required the lowest plant density followed by Kalamazoo, then East Lansing. For the Y95% level, the average plant density for Tuscola was 17.1 plants m^{-2}, Kalamazoo was 22.0 plants, m^{-2}, and East Lansing was 26.2 plants m^{-2}. Average plant density required to obtain Y95% in Michigan was 21.3 plants m^{-2} and the required density to obtain Y99% was 32.6 plants m^{-2}. Of the eight MI environments, three required different seeding rates for the different maturing cultivars. Adapted cultivars at Tuscola in 2009 and 2010 required greater plant densities compared to the earlier cultivar. At Kalamazoo in 2009, similar to the MN environments, the early cultivar required a greater seeding rate to attain optimum yield.

Iowa

Optimum plant populations for Iowa environments were found to be greater for the northerly environments relative to the southern, Corning, IA, environments (Table 2).

Average plant density required to achieve Y95% for Hudson was 24.0 plants m^{-2}, 31.5 plants m^{-2} for Story City, and 19.5 plants m^{-2} at Corning. Variability in the optimum plant density was observed across cultivar by environment combinations with values ranging from 4.4 plants m^{-2} to 39.1 plants m^{-2}. Of the 18 environment x cultivar combinations in IA, 13 required plant densities greater than 16.7 plants m^{-2}, and 8 required plant densities greater than 35.4 plants m^{-2}. Of the nine Iowa site-years, at two, the two different cultivars required different plant densities for optimal yield. At Hudson in 2011, the adapted cultivar required a stand of 4.4 plants m^{-2} compared to 16.7 plants m^{-2} for the early cultivar while at Corning in 2009, the adapted cultivar required a greater plant density (17.1 plants m^{-2}) than the earlier cultivar (9.9 plants m^{-2}) to obtain the same yield (5.14 Mg ha^{-1}).

Kentucky

Of the eight environments evaluated in KY from 2009 to 2011, the two cultivars had similar responsiveness to plant density, and in only one case did the two cultivars differ in their asymptotic yield (Lexington, 2009) where the later maturing cultivar yielded greater than the earlier cultivar. However, two of the eight KY environments were un-responsive to seeding rate; β parameters for New Haven 2009 and 2011 did not converge for Eq.1. The plant density to reach the 95% yield level in KY ranged from 13.1 plants m^{-2} at Hopkinsville to 40.5 plants m^{-2} at New Haven in 2010 with an overall average of 21.2 plants m^{-2} to obtain the Y95% level and 32 plants m^{-2} to reach the Y99% yield threshold (Table 2).

Arkansas

The plant density required to achieve 95% of the asymptotic yield in Arkansas environments range from 10 plants m^{-2} to 120 plants m^{-2} (Table 2). The 120 plants m^{-2} was required to achieve the Y95% threshold at Keiser in 2009 was beyond the highest seeding rate evaluated indicating a near linear, small β parameter, increase in yield across all plant populations evaluated. As a result the α parameter, yield asymptote, for this location was 7.49 Mg ha^{-1}, which well beyond the observed yield value. The highest treatment mean yield was approximately 5.2 Mg ha^{-1}. Based on the fitted equation for this environment, the predicted plant population required to achieve the 5.2 Mg ha^{-1} observed yield was approximately 54 plants m^{-2} (data not shown). This value is similar to the predicted optimum plant densities for Keiser in other years. Optimum plant densities varied by cultivar within environments. Keiser in 2010, Weiner in 2010 and 2011, and Colt in 2010 all had significant cultivar differences for β. The anticipated result of plant density of the later maturing cultivar being less than that required by the earlier maturing cultivar was observed at the Keiser and Weiner locations. However, this was not the case at Colt, where the earlier maturing cultivar required a lower plant population to obtain optimum yields compared to the later maturing cultivar. On average, excluding 2009 Keiser, the average plant density required to achieve obtain 95% or 99% of maximum yield, was greater for the Keiser environment with a requisite 39.6 plants m^{-2} for Y95% and 60.9 plants m^{-2} for Y99%. The Weiner location required 25.6 plants m^{-2} for 95% yield and 39.4 plants m^{-2} to reach the Y99% threshold. Lastly, the Colt location required the

lowest plant density with 14.7 plants m^{-2} needed for the Y95% threshold and 22.6 plants m^{-2} required for the Y99% level.

Louisiana

Optimum plant densities ranged from 7 plants m^{-2} to 16.6 plants m^{-2} (Table 2). Responsiveness to plant density did not differ for the two cultivars within a given location. Optimum seeding rates were generally more stable than other states. In fact, seven of ten environments had optimum seeding rates between 9.2 – 10.7 plants m^{-2}. Overall average seeding rate to obtain Y95% across Louisiana environments was 11.4 plants m^{-2} and 17.6 plants m^{-2} to reach the Y99% level.

Stand mortality

The overall analysis of covariance indicated that for each of the three regression measures of stand mortality (i.e. seeded vs emerged, seeded vs harvested, and emerged vs harvested) the main effects of cultivar maturity and the cultivar maturity by seeding density interaction were not significant. This finding indicates that stand establishment between the two cultivars was similar across all environments. For each mortality measure, an interaction between environment and seeding density was present which indicated differential stand establishment across the site-years ($P<0.01$). The percentage of total random variation associated with the random effects in each of the three analyses totaled approximately 20%, with the majority attributable to environmental differences. Because cultivar maturity and seeding density by maturity interactions were not significant, they were removed from the model which was re-fit. Separate regressions for each environment were fit to assess stand establishment variability across individual sites

(Table 4). The regression coefficients noted as β_1 in Table 4 is the indicator of stand loss. Values closest to 1 indicate small loss between plating, emergence, and harvest while values nearing 0 indicate greater loss. On average, the greatest loss in stand occurred from seeding until emergence which is indicated by an average regression coefficient β_1 of 0.81. Of the plants that germinated and emerged, 89% survived until harvest (β_1= 0.89). Cumulative stand losses from planting until harvest was similar to the product of the seeded: emerged and emerged: harvested regression coefficients with a value of β_1=0.74. By-environment regression coefficients revealed great variability in stand loss across all environments as well as considerable variation within a single location over years (Table 4). While characterizing seeding to emergence losses and emergence to harvest losses are important for understanding early season plant establishment issues and in-season stand attrition, the cumulative loss in stand from seeding until harvest integrates all phases of stand mortality which was used to determine optimum seed density required to achieve target harvested plant density. Each environment-specific seeded density: harvested density regression coefficient was applied to the optimum final plant density values in Table 2 to estimate the seeded density required to achieve 95% and 99% of asymptotic yield. This was accomplished by dividing the targeted final plant stand by the seed density: harvested density β_1 regression coefficient. Ultimately, when yield response to final plant stand is understood AND seeded density to harvested plant density mortality is established, optimum seeding rate rates for a given environment or environment by cultivar maturity combination can be estimated.

Conclusions

Results from this study indicate that full-season cultivars outperform short-season cultivars. This is not necessarily a new finding, but rather emphasizes that in most environments when soybean plants are able to capture additional resources (e.g. PAR, water, nutrients, etc.) and establish more vegetative biomass during growth and development, greater yield will be realized. Results based on the analysis of variance indicate that yield of full-season varieties seeded at a density of 30.9 seeds m^{-2} were not significantly different than the same cultivar seeded at higher rates. Although full-season cultivars were on average higher yielding, short season cultivars seeded at high densities were found to yield similarly to full-season cultivars, as was noted by Edwards and Purcell (2005).

Results from the non-linear modeling of the yield response to harvested plant density indicate much variation across environments for the plant densities required to achieve the Y95% and Y99% level of asymptotic yield. While latitude affected the yield response to plant population, its contribution to the overall variability was very small. A smaller harvested plant density was required in the south compared to the north, but inconsistency exists for within a given location over years, as well as within a narrow latitude range (i.e. within a given state). Figure 7 displays a histogram representing the distribution of seeding rates required to achieve Y95% for each site-year by cultivar combinations. At over 50% of the site-year-cultivar combinations, the harvested plant density that achieves Y95% was between 15 and 25 plants m^{-2}. As discussed stand loss/mortality measures coupled with the optimum harvest plant stand determinations

allows for indirect determination of the optimum seeding density. Given the variability observed in both stand loss coefficients and seeding density to reach a threshold yield level, the calculated variation in optimum seeded densities is complex. The optimum seeding rate target (i.e. Y95%-Y99%) for a producer will depend on the yield potential of the environment, costs associated with seed, and a commodity price expectation. When the yield goal is increased from Y95% to Y99%, the increase in optimum density at Y99% is 53% greater than Y95%; thus, different production circumstances may lead to different optimum seeding densities depending on producers' unique situations. Further research is needed to discern specific environment conditions contributing to variability for optimum harvested plant stands and stand attrition so that predictions and/or seeding rate recommendations can be made on a site-specific basis rather than the average result over a period of time within a particular geography.

Table 1. Details for experimental sites including year, state (ST), location name, latitude (lat.), longitude (long.), elevation (elev.), soil type, previous crop, previous fall tillage, spring tillage operation prior to planting, planting date, harvest date, and cultivars evaluated.

Year	ST	Location	Lat.	Long.	Elev.	Soil Type	Previous crop	† Fall Tillage	† Spring Tillage	Planting Date	Harvest Date	Short-season cultivar	Full-season cultivar
2009	MN	Crookston	47.8	-96.6	265	Wheatville loam	Wheat	CT	FC	21-May	29-Sep	90A06	90M60
2009	MN	Morris	45.6	-95.9	345	McIntosh silt loam	Corn	MP	FC	19-May	30-Sep	90Y41	91Y20
2009	MN	Becker	45.4	-93.9	289	Hubbard coarse loam	Rye	NT	FC	4-May	18-Oct	91Y20	92Y30
2009	MN	St. Paul	45.0	-93.2	290	Waukegan silt loam	Corn	MP	FC	6-May	16-Oct	91Y20	92Y30
2009	MN	Rosemount	44.7	-93.1	290	Waukegan silt loam	Corn	MP	FC	7-May	19-Oct	91Y20	92Y30
2009	MN	Waseca	44.1	-93.5	350	Webster clay loam	Corn	CT	FC	12-May	21-Oct	91Y20	92Y30
2009	MI	Tuscola	43.5	-83.7	191	Tappan-lando loam	Corn	CT	FC	19-May	23-Oct	91M01	91Y70
2009	MI	East Lansing	42.7	-84.5	261	Capac loam	Corn	CT	FC	1-Jun	21-Oct	91Y70	92Y80
2009	IA	Hudson	42.4	-92.5	274	Nevin silty clay	Corn	CT	FC	11-May	13-Oct	91Y70	92M54
2009	MI	Kalamazoo	42.3	-85.6	241	Fox sandy loam	Corn	CT	FC	23-May	22-Oct	91Y70	92Y80
2009	IA	Story City	42.2	-93.6	310	Kossuth silty clay loam	Corn	DK	FC	19-May	11-Oct	91Y70	92M54
2009	IA	Corning	41.0	-94.7	364	Macksburg silty clay loam	Corn	NT	NT	20-May	20-Oct	92M54	93M42
2009	KY	Lexington	38.0	-84.5	299	Mercer Silt Loam	Corn	NT	NT	18-May	11-Nov	93Y20	94Y01
2009	KY	New Haven	37.8	-85.7	140	Lindside Silt Loam	Corn	NT	NT	19-May	5-Nov	93Y20	94Y01
2009	KY	Hopkinsville	36.9	-87.5	162	Pembroke Silt Loam	Corn	NT	DK	20-May	26-Oct	93Y20	94Y01
2009	AR	Keiser	35.7	-90.1	71	Sharkey silty clay	Corn	DK	FC	23-Jun	21-Oct	93Y70	94Y70
2009	AR	Weiner	35.6	-90.9	75	Henry silt loam	Soybean	DK	NT	9-Jun	21-Oct	93Y70	94Y70
2009	LA	Winnsboro	32.2	-91.7	25	Gigger silt loam	Cotton	NT	BD	23-Apr	10-Sep	94M50	95M30
2009	LA	St. Joseph	31.9	-91.2	25	Sharkey clay	Sorghum	DK	FC	19-May	8-Oct	94Y70	95Y40
2010	MN	Crookston	47.8	-96.6	265	Wheatville loam	Barley	FC	FC	19-May	4-Oct	90A06	90M60
2010	MN	Becker	45.4	-93.9	289	Hubbard Loamy Sand	Rye	DK	FC	27-Apr	12-Oct	91Y20	92Y30
2010	MN	St. Paul	45.0	-93.2	290	Waukegan Silt Loam	Corn Silage	MP	FC	2-Jun	14-Oct	91Y20	92Y30
2010	MN	Rosemount	44.7	-93.1	290	Waukegan silt loam	Corn	CT	FC	19-May	6-Oct	91Y20	92Y30
2010	MN	Waseca	44.1	-93.5	350	Webster Clay Loam	Corn	CT	FC	4-May	4-Oct	91Y20	92Y30
2010	MI	Reese	43.5	-83.7	191	Tappan-Lando loam	Corn	CT	FC	6-May	27-Sep	91M01	91Y70

2010	IA	Hudson	42.4	-92.5	274	Dinsmore silty clay loam	Corn	CT	FC	6-May	4-Oct	91Y70	92M54
2010	MI	Kalamazoo	42.3	-85.6	241	Kalamazoo loam	Corn	CT	FC	6-May	30-Sep	91Y70	92Y80
2010	IA	Story City	42.2	-93.6	310	Nicollet loam	Corn	NT	DK	10-May	5-Oct	91Y70	92M54
2010	IA	Corning	41.0	-94.7	364	Macksburg silty clay loam	Corn	NT	NT	3-May	6-Oct	92M54	93M42
2010	KY	New Haven	38.0	-84.5	299	Lindside Silt Loam	Corn	NT	NT	8-Jun	11-Oct	93Y20	94Y01
2010	KY	Lexington	37.8	-85.7	140	Maury silt loam	Corn	NT	NT	2-Jun	6-Oct	93Y20	94Y01
2010	AR	Weiner	35.7	-90.1	71	Henry silt loam	Soybean	DK	NT	8-May	13-Oct	93Y70	94Y70
2010	AR	Colt	35.6	-90.9	75	Calloway Silt Loam	Rice	DK	FC	21-May	1-Oct	93Y70	94Y70
2010	AR	Keiser	35.1	-90.7	110	Sharkey silty clay	Corn	DK	FC	8-May	10-Oct	93Y70	94Y70
2010	LA	Winnsboro	32.2	-91.7	25	Gigger silt loam	Cotton	NT	BD	4-May	13-Sep	94Y70	95Y40
2010	LA	St. Joseph	31.9	-91.2	25	Sharkey Clay	Sorghum	DK	BD	3-May	23-Sep	94Y70	95Y40
2010	LA	Baton Rouge	30.5	-91.1	26	Commerce silt loam	Soybean	NT	DK	19-Apr	8-Sep	94Y70	95Y40
2010	LA	Crowley	30.2	-92.4	8	Crowley silt loam	Fallow	NT	DK	21-May	27-Sep	94Y70	95Y40
2011	MN	Crookston	47.8	-96.6	265	Wheatville loam	Wheat	CT	FC	21-May	6-Oct	90A06	90M60
2011	MN	Becker	45.4	-93.9	289	Hubbard Loamy Sand	Rye	DK	FC	3-May	3-Oct	91Y20	92Y30
2011	MN	St. Paul	45.0	-93.2	290	Waukegan Silt Loam	Corn	MP	FC	4-May	11-Oct	91Y20	92Y30
2011	MN	Rosemount	44.7	-93.1	290	Waukegan silt loam	Corn	CT	FC	1-Jun	4-Oct	91Y20	92Y30
2011	MN	Waseca	44.1	-93.5	350	Webster Clay Loam	Corn	CT	FC	18-May	5-Oct	91Y20	92Y30
2011	MI	Reese	43.5	-83.7	191	Tappan-Lando loam	Corn	CT	FC	17-May	12-Oct	91M01	91Y70
2011	MI	East Lansing	42.7	-84.5	261	Capac loam	Corn	NT	NT	4-Jun	21-Oct	91Y70	92Y80
2011	IA	Hudson	42.4	-92.5	274	Dinsmore silty clay loam	Corn	NT	FC	2-Jun	17-Oct	91Y70	92M54
2011	MI	Kalamazoo	42.3	-85.6	241	Kalamazoo loam	Corn	CT	DK	10-May	29-Oct	91Y70	92Y80
2011	IA	Story City	42.2	-93.6	310	Kossuth silty clay loam	Corn	NT	DK	6-May	6-Oct	91Y70	92M54
2011	IA	Corning	41.0	-94.7	364	Macksburg silty clay loam	Corn	NT	NT	3-May	14-Oct	92M54	93M42
2011	LA	New Haven	38.0	-84.5	299	Lindside Silt Loam	Corn	NT	NT	8-Jun	25-Oct	93Y20	94Y01
2011	LA	Lexington	37.8	-85.7	140	Maury silt loam	Corn	NT	NT	30-May	11-Oct	93Y20	94Y01
2011	LA	Hopkinsville	36.9	-87.5	162	Pembroke Silt Loam	Corn	NT	NT	31-May	10-Oct	93Y20	94Y01
2011	AR	Weiner	35.7	-90.1	71	Henry silt loam	Rice	NT	DK	9-Jun	27-Oct	93Y70	94Y70
2011	AR	Colt	35.6	-90.9	75	Calloway Silt Loam	Rice	DK	FC	5-Jun	17-Oct	93Y70	94Y70
2011	AR	Keiser	35.1	-90.7	110	Sharkey silty clay	Corn	DK	FC	13-Jun	1-Nov	93Y70	94Y70
2011	LA	Winnsboro	32.2	-91.7	25	Gigger silt loam	Cotton	NT	BD	14-Apr	31-Aug	94Y70	95Y40
2011	LA	St. Joseph	31.9	-91.2	25	Sharkey Clay	Sorghum	DK	BD	11-May	27-Sep	94Y70	95Y40
2011	LA	Baton Rouge	30.5	-91.1	26	Commerce silt loam	Soybean	NT	DK	21-Apr	26-Aug	94Y70	95Y40

2011	LA	Crowley	30.2	-92.4	8	Crowley silt loam	Fallow/SB	NT	DK	11-May	5-Oct	94Y70	95Y40

† CT-conservation tillage, MP-moldboard plow, DK-disk, NT-no tillage, FC- field cultivator, BD-raised bed

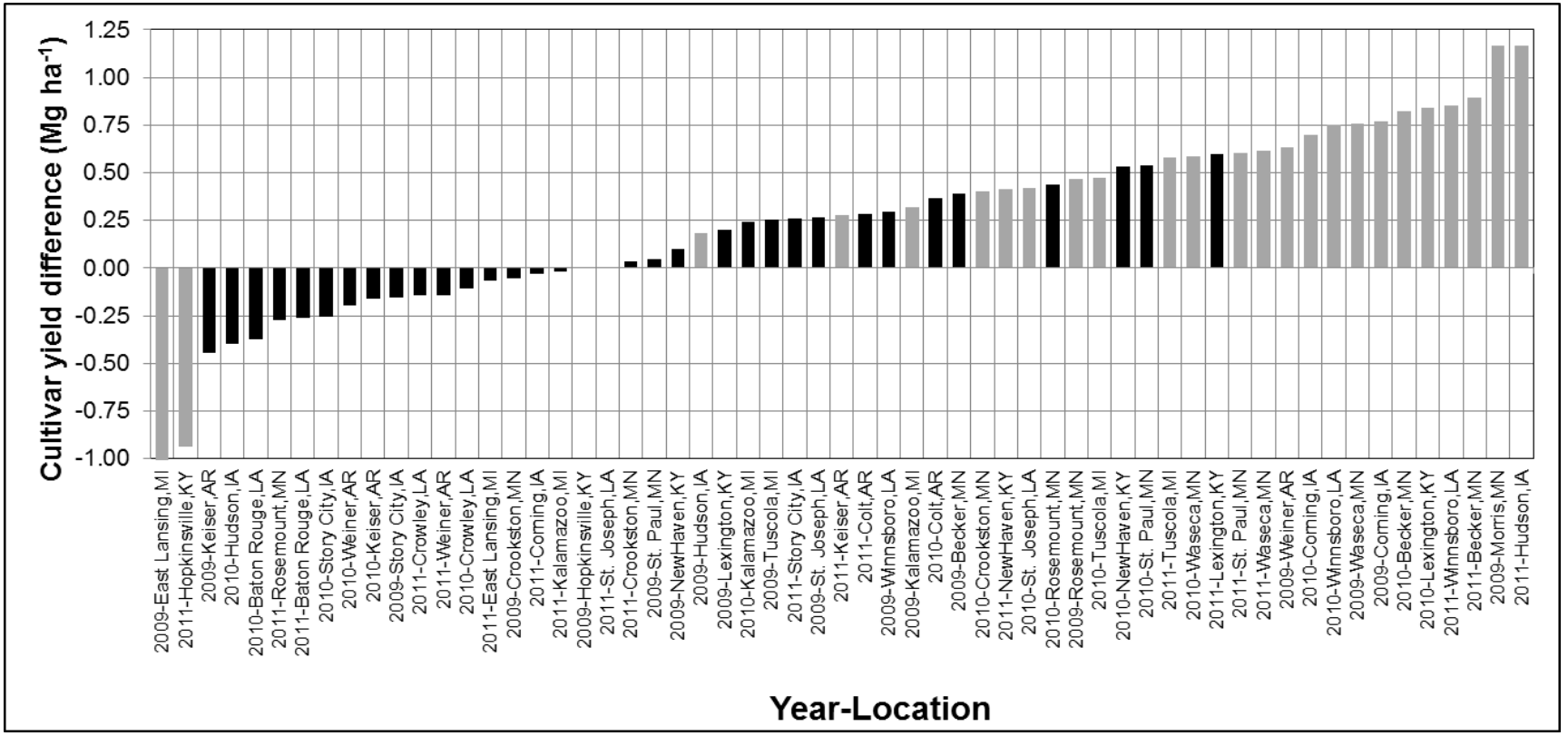


Figure 1. Deviation for yield (Mg ha^{-1}) between full-season and short-season soybean cultivars across years and locations. Yield differences were significantly different (P<0.05) at yield-locations gray shaded. Full-season cultivars outperformed short-season cultivars, numerically, 41 of the 59 environments (70% of the time). Of the 25 year/location combinations that had significant differences, at 23 the full-season variety significantly outperformed the short-season variety (92% of the time).

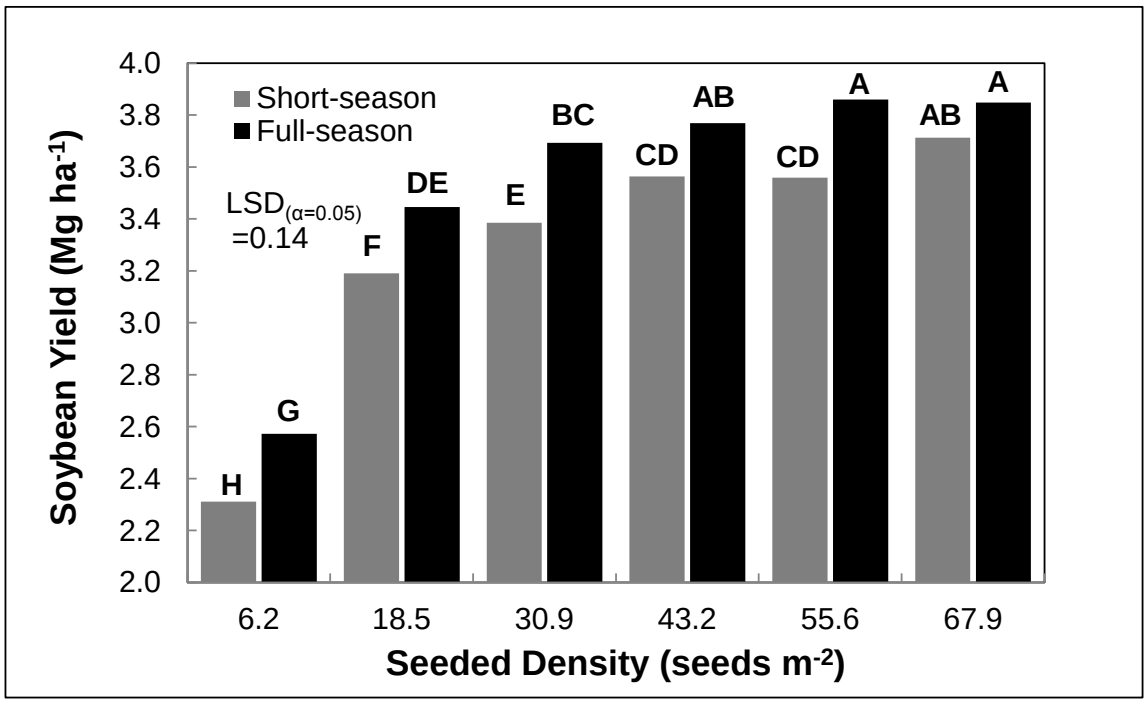


Figure 2. Yield for short- and full-season soybean cultivars planted at differing seeding densities. Data collected from 59 site-years of data across United States environments.

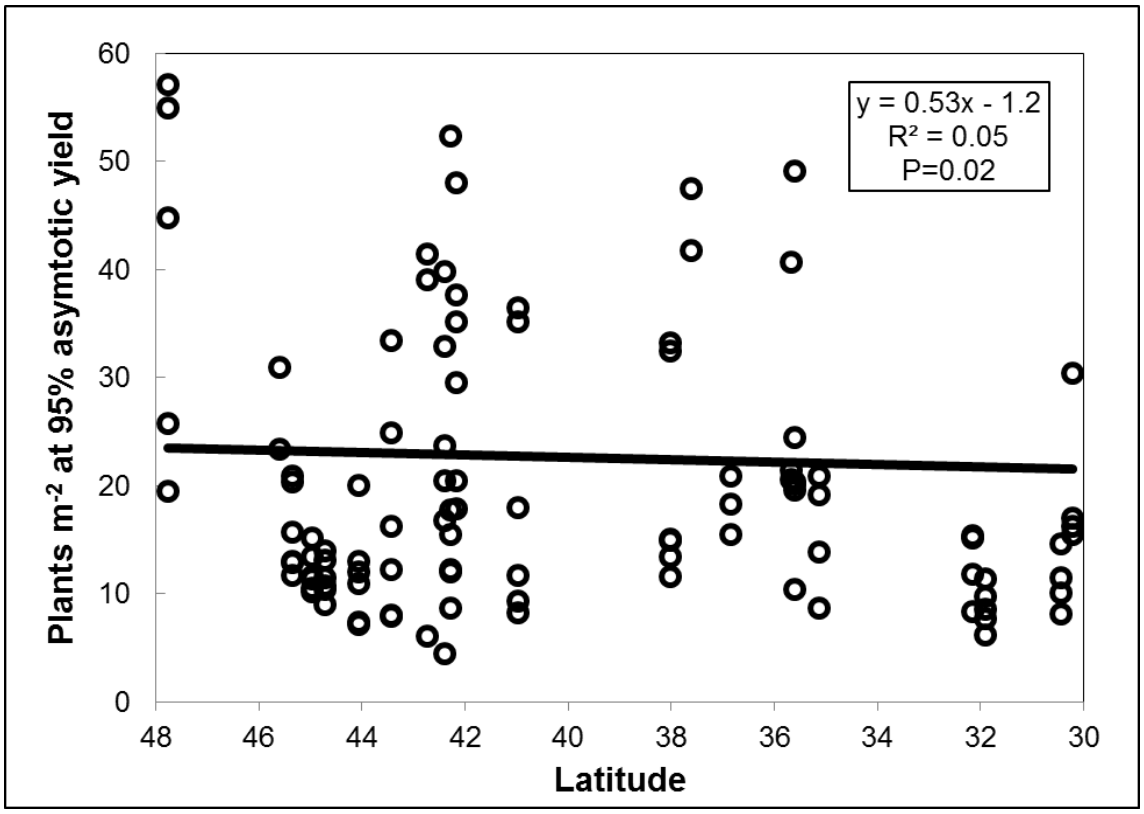


Figure 3. Plant population (plants m^{-2}) at 95% asymptotic yield for each site-year by cultivar combination as determined by Eq. 1 regressed against site-year latitude. Three location by cultivar combinations were excluded from this analysis which were grown at Keiser, AR because plant densities required to achieve 95% yield were greater than 95 plants m^{-2}., which substantially deviated from data obtained from all other locations.

Table 2. Reduced model parameters fit from Eq.1 for each location. Information incudes year, location, latitude, cultivar maturity, α (asymptote) and β (responsiveness) parameters with their corresponding 95% confidence intervals, pseudo-R^2 for each location model, the R8 plant density (PD) required to achieve 95% and 99% of the asymptotic (α) yield, and the seeding density (SD) required to reach 95% and 99% asymptotic yield based on the seeded: harvested stand loss regression coefficients from Table 4.

Year	Location	Maturity	α	α Low 95% CI	α Upper 95% CI	β	β Lower 95% CI	β Upper 95% CI	Pseudo-R^2	PD at 95% asymptote	PD at 99% asymptote	SD at 95% asymptote	SD at 99% asymptote
2009	Crookston	Early	1.825	1.622	2.029	0.0053	0.0046	0.0061	0.92	56.1	86.2	101.8	156.0
2009	Crookston	Adapted	2.937	2.676	3.197	§	§	§	-----	¶	¶	††	††
2010	Crookston	†	3.200	3.099	3.302	0.0136	0.0111	0.0161	-----	22.0	33.9	25.0	36.9
2011	Crookston	†	3.804	3.652	3.955	0.0053	0.0046	0.0061	-----	56.1	86.2	63.1	92.6
2009	Morris	Early	2.276	2.088	2.465	0.0107	0.0077	0.0137	0.58	28.0	43.0	48.0	70.5
2009	Morris	Adapted	2.818	2.610	3.026	§	§	§	-----	¶	¶	††	††
2009	Becker	†	4.509	4.382	4.636	0.0169	0.0133	0.0206	0.58	17.7	27.2	20.6	29.5
2010	Becker	Early	4.174	4.004	4.344	0.0147	0.0113	0.0181	-----	20.4	31.3	23.9	35.8
2010	Becker	Adapted	4.649	4.446	4.851	0.0232	0.0162	0.0301	-----	12.9	19.8	15.8	23.3
2011	Becker	†	3.995	3.872	4.119	0.0244	0.0185	0.0302	-----	12.3	18.9	16.0	23.1
2009	St. Paul	Early	3.427	3.211	3.643	0.0251	0.0204	0.0298	0.39	11.9	18.3	24.7	35.1
2009	St. Paul	Adapted	4.105	3.902	4.308	§	§	§	-----	¶	¶	††	††
2010	St. Paul	Early	3.494	3.291	3.697	§	§	§	-----	¶	¶	15.7	22.6
2010	St. Paul	Adapted	4.127	3.928	4.325	§	§	§	-----	¶	¶	††	††
2011	St. Paul	Early	3.748	3.550	3.945	§	§	§	-----	¶	¶	13.8	20.2
2011	St. Paul	Adapted	3.485	3.283	3.687	§	§	§	-----	¶	¶	††	††
2009	Rosemount	Early	3.553	3.409	3.698	0.0271	0.0237	0.0306	0.81	11.1	17.0	9.3	14.7
2009	Rosemount	Adapted	4.196	4.057	4.334	§	§	§	-----	¶	¶	††	††
2010	Rosemount	Early	3.233	3.098	3.368	§	§	§	-----	¶	¶	17.8	25.1
2010	Rosemount	Adapted	4.084	3.949	4.219	§	§	§	-----	¶	¶	††	††
2011	Rosemount	Early	2.896	2.760	3.033	§	§	§	-----	¶	¶	16.0	23.3
2011	Rosemount	Adapted	2.462	2.326	2.597	§	§	§	-----	¶	¶	††	††
2009	Waseca	Early	3.494	3.267	3.720	0.0149	0.0097	0.0202	0.61	20.1	30.9	33.1	47.8
2009	Waseca	Adapted	4.109	3.923	4.294	0.0400	0.0193	0.0608	-----	7.5	11.5	15.8	21.3
2010	Waseca	†	4.660	4.532	4.788	0.0268	0.0223	0.0312	-----	11.2	17.2	15.5	22.1
2011	Waseca	†	4.169	4.038	4.300	§	§	§	-----	¶	¶	16.1	24.1
2009	Tuscola	Early	3.372	3.264	3.480	0.0190	0.0150	0.0230	0.80	15.8	24.2	18.9	29.0
2009	Tuscola	Adapted	‡	‡	‡	0.0112	0.0088	0.0136	-----	26.7	41.1	32.0	49.1
2010	Tuscola	Early	2.821	2.729	2.913	0.0224	0.0121	0.0327	-----	13.4	20.6	16.5	22.4
2010	Tuscola	Adapted	‡	‡	‡	0.0097	0.0073	0.0121	-----	30.9	47.5	30.8	44.4
2011	Tuscola	Early	3.458	3.333	3.583	0.0374	0.0280	0.0468	-----	8.0	12.3	14.3	20.9
2011	Tuscola	Adapted	3.785	3.670	3.899	§	§	§	-----	¶	¶	††	††
2009	East Lansing	†	2.968	2.648	3.288	0.0069	0.0045	0.0094	0.68	43.4	66.7	57.6	87.0
2011	East Lansing	†	3.576	3.415	3.736	0.0334	0.0233	0.0435	-----	9.0	13.8	6.7	11.0

2009	Hudson	†	4.560	4.336	4.784	0.0134	0.0102	0.0165	0.74	22.4	34.4	40.5	59.8
2010	Hudson	Early	4.661	4.467	4.854	0.0077	0.0065	0.0089	-----	39.1	60.0	63.8	94.4
2010	Hudson	Adapted	5.412	5.232	5.592	§	§	§	-----	¶	¶	††	††
2011	Hudson	Early	4.661	4.467	4.854	0.0179	0.0115	0.0244	-----	16.7	25.7	36.1	50.7
2011	Hudson	Adapted	5.412	5.232	5.592	0.0684	0.0329	0.1040	-----	4.4	6.7	15.9	19.8
2009	Kalamazoo	Early	2.491	2.283	2.699	0.0045	0.0029	0.0060	0.89	66.7	102.6	68.8	106.1
2009	Kalamazoo	Adapted	‡	‡	‡	0.0174	0.0093	0.0254	-----	17.2	26.5	17.2	26.8
2010	Kalamazoo	Early	4.369	4.187	4.552	0.0250	0.0213	0.0287	-----	12.0	18.4	17.4	25.8
2010	Kalamazoo	Adapted	5.214	5.025	5.402	§	§	§	-----	¶	¶	17.4	25.8
2011	Kalamazoo	Early	3.249	3.065	3.433	§	§	§	-----	¶	¶	15.2	23.4
2011	Kalamazoo	Adapted	4.020	3.841	4.199	§	§	§	-----	¶	¶	††	††
2009	Story City	Early	3.465	3.160	3.770	0.0157	0.0104	0.0210	0.66	19.1	29.3	31.1	45.6
2009	Story City	Adapted	4.026	3.720	4.332	§	§	§	-----	¶	¶	††	††
2010	Story City	†	3.879	3.576	4.181	0.0080	0.0059	0.0100	-----	37.7	57.9	59.4	89.0
2011	Story City	Early	2.858	2.533	3.182	§	§	§	-----	¶	¶	50.4	77.2
2011	Story City	Adapted	3.405	3.034	3.777	§	§	§	-----	¶	¶	††	††
2009	Corning	Early	5.135	5.032	5.238	0.0304	0.0213	0.0395	0.72	9.9	15.1	30.8	41.1
2009	Corning	Adapted	‡	‡	‡	0.0175	0.0126	0.0224	-----	17.1	26.3	45.0	63.0
2010	Corning	†	‡	‡	‡	0.0085	0.0074	0.0096	-----	35.4	54.4	47.4	68.8
2011	Corning	†	4.558	4.424	4.692	0.0311	0.0245	0.0378	-----	9.6	14.8	24.7	33.7
2009	Lexington	Early	4.582	4.144	5.020	0.0091	0.0063	0.0119	0.66	33.0	50.7	46.3	69.6
2009	Lexington	Adapted	5.579	5.062	6.096	§	§	§	-----	¶	¶	46.3	69.6
2010	Lexington	†	3.125	2.884	3.366	0.0215	0.0154	0.0276	-----	13.9	21.4	††	††
2011	Lexington	†	3.982	3.732	4.233	§	§	§	-----	¶	¶	25.5	38.5
2009	New Haven	†	3.467	3.263	3.671	#	#	#	0.57	#	#	#	#
2010	New Haven	†	3.681	3.233	4.130	0.0074	0.0046	0.0102	-----	40.5	62.2	48.2	74.7
2011	New Haven	†	4.218	4.001	4.434	#	#	#	-----	#	#	#	#
2009	Hopkinsville	†	3.857	3.640	4.075	0.0228	0.0148	0.0308	0.50	13.1	20.2	34.3	45.9
2011	Hopkinsville	†	2.563	2.353	2.774	§	§	§	-----	¶	¶	28.6	39.8
2009	Keiser	†	7.489	5.019	9.959	0.0025	0.0013	0.0037	0.80	120.3	184.9	209.5	321.0
2010	Keiser	Early	4.050	3.476	4.625	0.0044	0.0027	0.0061	-----	67.6	104.0	107.1	161.3
2010	Keiser	Adapted	‡	‡	‡	0.0062	0.0034	0.0090	-----	48.5	74.5	78.5	117.4
2011	Keiser	Early	4.005	3.667	4.343	0.0142	0.0108	0.0177	-----	21.1	32.4	48.2	73.9
2011	Keiser	Adapted	4.858	4.443	5.273	§	§	§	-----	¶	¶	††	††
2009	Weiner	†	2.795	2.688	2.902	0.0150	0.0124	0.0177	0.76	20.0	30.7	39.1	58.5
2010	Weiner	Early	2.308	2.188	2.427	0.0053	0.0041	0.0066	-----	56.4	86.7	108.8	159.3
2010	Weiner	Adapted	‡	‡	‡	0.0125	0.0085	0.0166	-----	24.0	36.8	54.7	76.1
2011	Weiner	Early	2.653	2.559	2.747	0.0128	0.0098	0.0159	-----	23.4	36.0	45.7	67.8
2011	Weiner	Adapted	‡	‡	‡	0.0299	0.0217	0.0380	-----	10.0	15.4	22.1	31.6
2010	Colt	Early	1.713	1.553	1.874	0.0246	0.0160	0.0332	0.92	12.2	18.7	20.8	31.0
2010	Colt	Adapted	2.871	2.700	3.042	§	§	§	-----	¶	¶	††	††
2011	Colt	Early	4.477	4.347	4.606	0.0209	0.0162	0.0255	-----	14.3	22.0	26.3	38.3
2011	Colt	Adapted	‡	‡	‡	0.0148	0.0118	0.0178	-----	20.2	31.1	35.5	52.5
2009	Winnsboro	Early	3.768	3.542	3.995	0.0279	0.0220	0.0338	0.73	10.7	16.5	20.6	27.8

2009	Winnsboro	Adapted	4.198	3.970	4.425	§	§	§	-----	¶	¶	††	††
2010	Winnsboro	Early	4.067	3.851	4.284	§	§	§	-----	¶	¶	15.7	23.1
2010	Winnsboro	Adapted	3.669	3.459	3.878	§	§	§	-----	¶	¶	††	††
2011	Winnsboro	Early	2.616	2.401	2.830	§	§	§	-----	¶	¶	18.9	27.6
2011	Winnsboro	Adapted	2.157	1.947	2.367	§	§	§	-----	¶	¶	††	††
2009	St. Joseph	†	4.364	4.254	4.475	0.0423	0.0328	0.0519	0.89	7.1	10.9	15.3	22.1
2010	St. Joseph	†	5.248	5.139	5.357	0.0324	0.0283	0.0364	-----	9.2	14.2	19.4	28.9
2011	St. Joseph	Early	4.028	3.878	4.178	0.0262	0.0203	0.0320	-----	11.4	17.6	17.7	27.3
2011	St. Joseph	Adapted	2.830	2.704	2.956	#	#	#	-----	#	#	#	#
2010	Baton Rouge	†	4.615	4.421	4.808	0.0284	0.0185	0.0382	0.71	10.5	16.2	18.6	26.1
2011	Baton Rouge	†	2.760	2.571	2.949	§	§	§	-----	¶	¶	14.8	21.2
2010	Crowley	Early	4.579	4.426	4.733	0.0180	0.0151	0.0209	0.95	16.6	25.6	21.4	31.8
2010	Crowley	Adapted	3.488	3.336	3.639	§	§	§	-----	¶	¶	††	††
2011	Crowley	†	1.463	1.351	1.576	§	§	§	-----	¶	¶	††	††

† A single equation describes soybean yield response to final plant density for a particular environment.
‡ A single alpha describes soybean yield response to final plant density for a particular environment.
§ A single beta describes soybean yield response to final plant density for a particular environment.
¶ The calculated final plant stand required to achieve 95% and 99% of maximum yield is not different than the preceding values.
# Model did not converge for the parameter.
†† The calculated seeded density required to achieve 95% and 99% of maximum yield is not different than the preceding values.

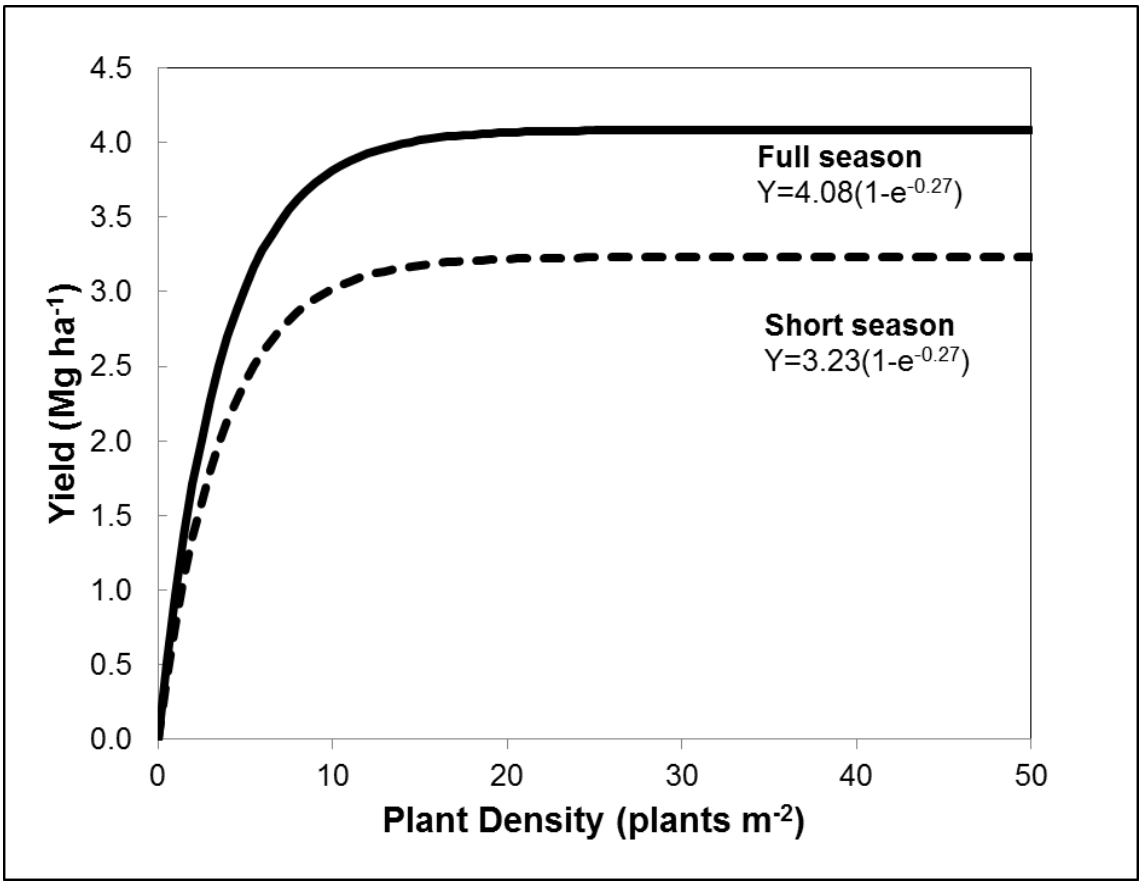


Figure 4. Parameter estimates determined for full-season and short-season cultivars grown at Rosemount, MN in 2010 from the reduced overall Rosemount Eq.1 model. Differences were observed for the asymptote (α) of full-season and short-season cultivars. β responsiveness parameters for both cultivars did not differ.

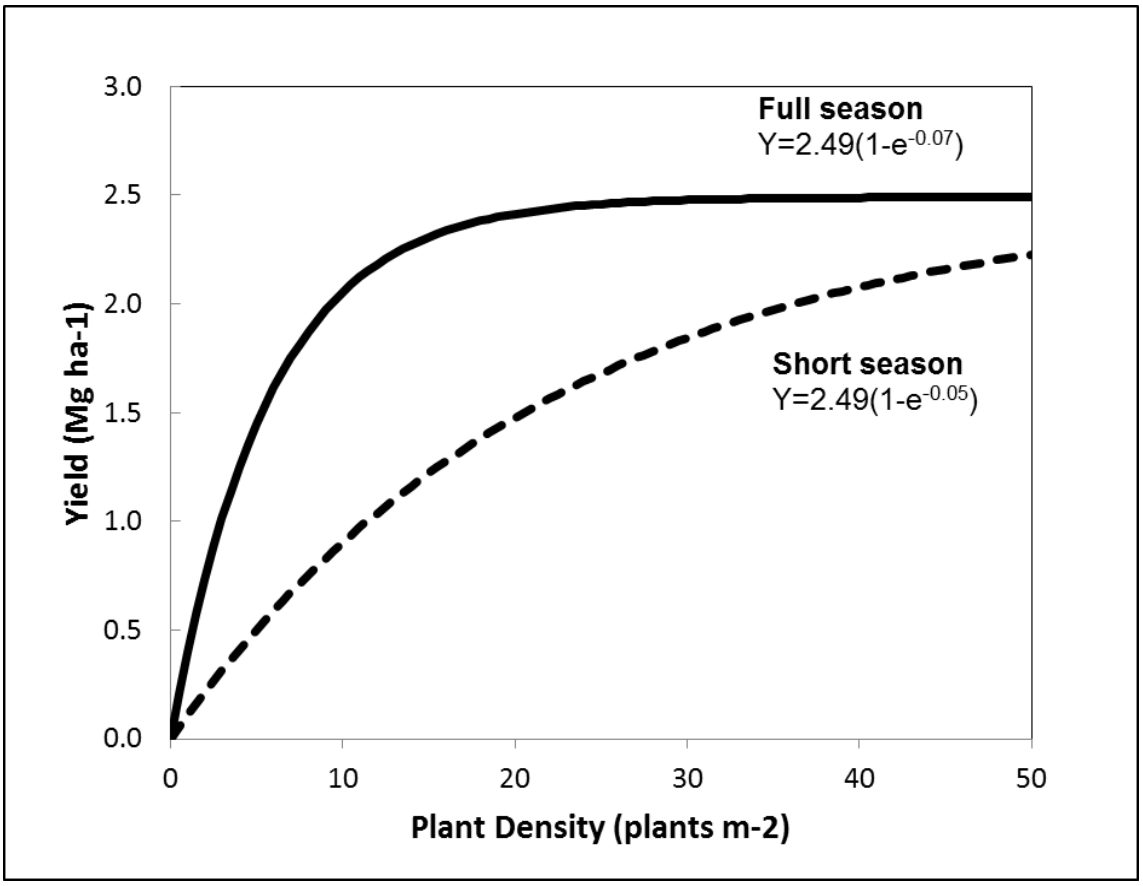


Figure 5. Parameter estimates determined for full-season and short-season cultivars grown at Kalamazoo, MI in 2009 from the reduced overall Kalamazoo Eq.1 model. Asymptotes (α) of full-season and short-season cultivars did not differ while responsiveness parameters (β) were different for each cultivar.

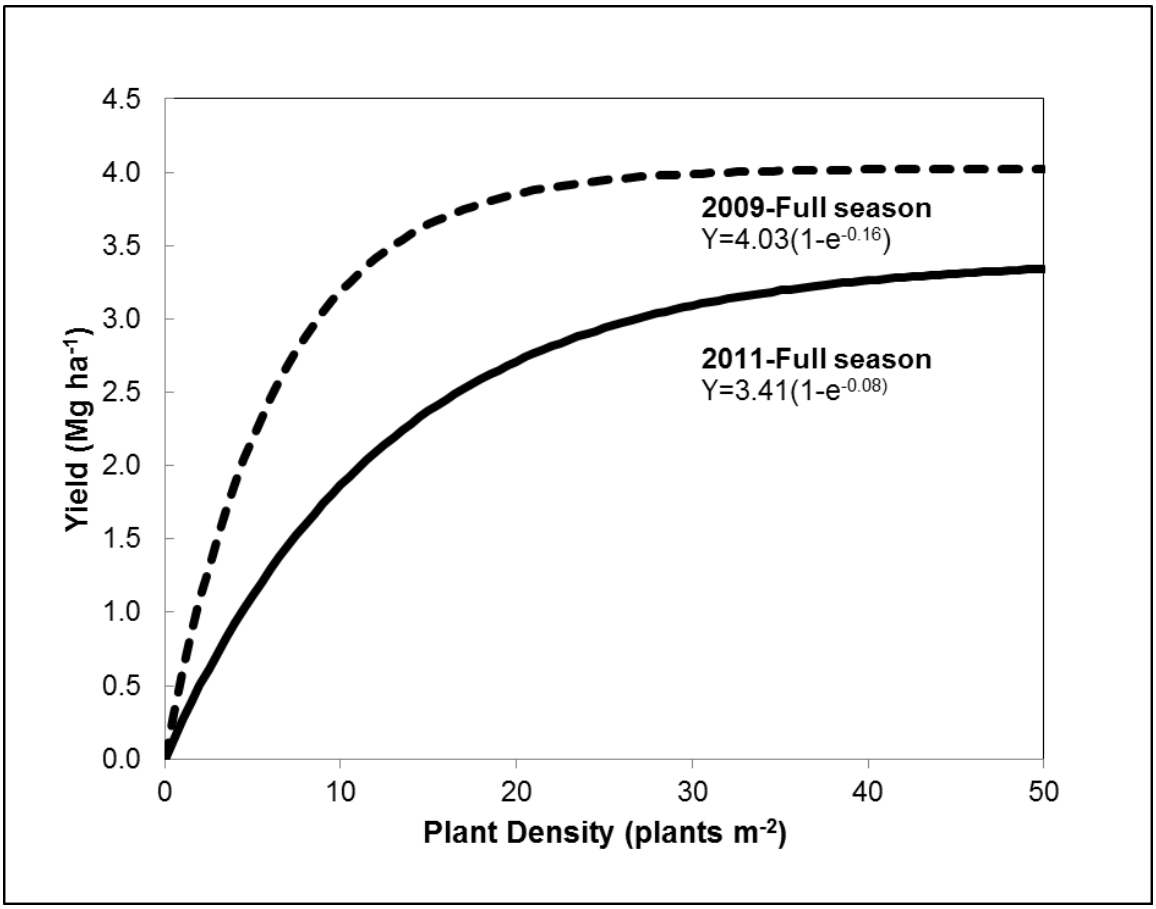


Figure 6. Parameter estimates determined for full-season cultivars grown at Story City, IA in 2009 and 2010 from the reduced overall Story City Eq.1 model. Differences were observed for the asymptote (α) and (β) responsiveness parameters for the same cultivar grown in different years.

Table 3. Results summary of the non-linear model parameter estimates. The number of environments where differences in α (asymptotic yield) and β (plant density responsiveness) are indicated in column 3. For environments where significant differences for either parameter were observed, columns 4 through 7 indicate the direction of the difference (short-season or full-season).

			Cultivar differences			
Parameter Testing Summary		**N=**	**Greater Yield (<α)**		**Greater PD Responsiveness (<β)**	
α	**β**	**environments**	**Full season**	**Short season**	**Full season**	**Short season**
No difference	No difference	24	------	------	------	------
--------	Difference	8	------	------	4	4
--------	No convergence	2	------	------	------	------
Difference	No difference	21	13	5	------	------
--------	Difference	3	3	0	0	3
--------	No convergence	1	0	1	------	------
Total		***59***	***16***	***6***	***4***	***7***

Table 4. Linear regression model parameters describing the soybean stand loss for three periods of time; from seeding density (SD) to emergence density (ED), SD to harvest density (HD), and ED to HD. Full-season and short-season cultivars did not differ significantly (P>0.05) for each stand loss measure, so pooled over individual location regression coefficients and R^2s are displayed. All location specific models were highly significant with P<0.001.

			ED = β1(SD)+ β0					HD = β1(ED)+ β0					HD = β1(SD)+ β0				
Year	**Location**	**ST**	**β_1**	**± *SE***	**β_0**	**± *SE***	**R^2**	**β_1**	**± *SE***	**β_0**	**± *SE***	**R^2**	**β_1**	**± *SE***	**β_0**	**± *SE***	**R^2**
2009	Crookston	MN	0.58	*0.02*	2.76	*0.96*	0.94	0.90	*0.05*	-0.24	*1.40*	0.86	0.56	*0.02*	0.99	*1.00*	0.91
2010	Crookston	MN	0.79	*0.04*	6.52	*1.58*	0.85	1.12	*0.08*	-0.53	*2.80*	0.69	0.99	*0.04*	2.77	*1.58*	0.93
2011	Crookston	MN	1.03	*0.05*	6.92	*2.26*	0.90	0.90	*0.06*	5.60	*2.64*	0.85	1.02	*0.04*	8.21	*1.67*	0.92
2009	Morris	MN	0.67	*0.03*	6.21	*1.11*	0.90	†	†	†	†	†	†	†	†	†	†
2009	Becker	MN	1.12	*0.04*	2.14	*1.58*	0.94	0.93	*0.03*	3.19	*1.42*	0.97	1.07	*0.03*	4.05	*1.20*	0.97
2010	Becker	MN	1.03	*0.03*	0.85	*1.45*	0.97	0.89	*0.02*	1.08	*0.87*	0.98	0.92	*0.03*	1.78	*1.59*	0.94
2011	Becker	MN	1.08	*0.02*	1.37	*0.96*	0.97	0.85	*0.02*	2.07	*1.15*	0.97	0.93	*0.02*	2.91	*1.27*	0.97
2009	St. Paul	MN	0.74	*0.03*	3.66	*1.53*	0.93	0.85	*0.03*	2.24	*0.92*	0.95	0.62	*0.03*	5.44	*1.60*	0.89
2010	St. Paul	MN	1.04	*0.03*	0.70	*1.37*	0.96	0.88	*0.03*	2.70	*1.51*	0.95	0.93	*0.03*	2.82	*1.82*	0.95
2011	St. Paul	MN	1.04	*0.02*	1.08	*0.96*	0.98	0.96	*0.02*	0.96	*1.34*	0.98	1.00	*0.03*	1.81	*1.32*	0.97
2009	Rosemount	MN	1.35	*0.03*	-4.37	*1.14*	0.98	0.79	*0.03*	3.12	*1.43*	0.93	1.09	*0.04*	-0.94	*1.38*	0.85
2010	Rosemount	MN	1.00	*0.04*	2.12	*1.66*	0.94	0.79	*0.03*	3.14	*1.15*	0.95	0.81	*0.03*	4.10	*1.17*	0.90
2011	Rosemount	MN	0.95	*0.04*	0.87	*1.59*	0.94	0.83	*0.03*	2.62	*1.28*	0.94	0.81	*0.03*	2.43	*1.17*	0.93
2009	Waseca	MN	1.00	*0.04*	2.63	*1.56*	0.91	0.74	*0.03*	3.39	*1.06*	0.94	0.73	*0.04*	5.53	*1.62*	0.84
2010	Waseca	MN	0.95	*0.03*	2.33	*1.20*	0.95	0.93	*0.03*	1.35	*1.06*	0.96	0.90	*0.03*	2.99	*1.25*	0.96
2011	Waseca	MN	0.93	*0.03*	-0.22	*1.35*	0.96	0.82	*0.04*	1.38	*1.68*	0.88	0.75	*0.05*	1.31	*2.15*	0.85
2009	Tuscola	MI	0.86	*0.02*	1.25	*1.13*	0.97	0.96	*0.05*	-0.66	*1.85*	0.89	0.84	*0.05*	0.19	*1.78*	0.86
2010	Tuscola	MI	0.86	*0.04*	1.42	*2.00*	0.90	1.29	*0.11*	8.09	*4.98*	0.78	1.22	*0.08*	5.51	*4.19*	0.85
2011	Tuscola	MI	0.65	*0.04*	1.97	*1.61*	0.84	†	†	†	†	†	†	†	†	†	†
2009	East Lansing	MI	0.95	*0.04*	2.48	*1.53*	0.94	0.85	*0.04*	0.28	*1.42*	0.91	0.79	*0.05*	2.90	*2.17*	0.85
2011	East Lansing	MI	0.79	*0.02*	2.59	*0.91*	0.97	1.39	*0.05*	-4.14	*2.20*	0.94	1.12	*0.04*	-1.33	*1.78*	0.94
2009	Hudson	IA	0.74	*0.04*	2.08	*1.37*	0.86	0.81	*0.07*	3.94	*2.02*	0.63	0.62	*0.05*	4.71	*1.86*	0.70
2010	Hudson	IA	0.81	*0.04*	6.53	*1.75*	0.91	0.82	*0.05*	2.28	*1.94*	0.87	0.69	*0.04*	6.84	*2.11*	0.86
2011	Hudson	IA	0.60	*0.04*	2.92	*1.98*	0.83	0.95	*0.06*	7.48	*1.81*	0.88	0.61	*0.05*	8.77	*1.92*	0.81
2009	Kalamazoo	MI	0.96	*0.02*	-0.73	*0.69*	0.99	†	†	†	†	†	†	†	†	†	†
2010	Kalamazoo	MI	0.87	*0.02*	3.01	*1.60*	0.98	0.86	*0.04*	-0.29	*2.04*	0.92	0.76	*0.03*	1.74	*1.32*	0.92
2011	Kalamazoo	MI	0.81	*0.04*	0.51	*1.86*	0.92	0.94	*0.05*	0.89	*2.06*	0.88	0.79	*0.04*	0.10	*2.09*	0.91
2009	Story City	IA	0.81	*0.05*	0.77	*3.59*	0.90	0.79	*0.06*	5.61	*2.84*	0.85	0.71	*0.04*	4.07	*1.54*	0.85

2010	Story City	IA	0.80	*0.03*	6.03	*1.70*	0.94	0.84	*0.03*	-0.46	*1.44*	0.95	0.68	*0.04*	4.35	*1.48*	0.89
2011	Story City	IA	0.86	*0.03*	-2.70	*1.11*	0.29	0.89	*0.04*	2.16	*1.57*	0.93	0.76	*0.04*	0.59	*1.89*	0.87
2009	Corning	IA	0.56	*0.04*	7.08	*1.63*	0.83	0.83	*0.09*	7.70	*2.73*	0.69	0.51	*0.05*	11.50	*2.09*	0.66
2010	Corning	IA	0.95	*0.05*	7.13	*2.15*	0.88	0.91	*0.03*	2.09	*2.16*	0.95	0.89	*0.05*	7.68	*2.84*	0.89
2011	Corning	IA	0.83	*0.07*	6.18	*2.99*	0.74	0.63	*0.05*	5.52	*1.77*	0.62	0.57	*0.06*	7.91	*2.23*	0.68
2009	Lexington	KY	0.79	*0.03*	-0.32	*1.81*	0.95	0.89	*0.09*	5.09	*2.54*	0.74	0.76	*0.06*	2.88	*2.37*	0.80
2010	Lexington	KY	0.37	*0.03*	2.12	*1.53*	0.79	1.03	*0.07*	2.69	*1.32*	0.83	0.42	*0.03*	3.22	*1.58*	0.78
2011	Lexington	KY	0.67	*0.07*	1.02	*2.95*	0.71	0.72	*0.06*	3.52	*1.60*	0.79	0.58	*0.06*	1.43	*2.16*	0.32
2010	New Haven	KY	0.49	*0.03*	1.80	*1.65*	0.83	0.64	*0.10*	6.73	*2.09*	0.47	0.31	*0.06*	7.94	*2.36*	0.41
2009	New Haven	KY	0.76	*0.03*	1.78	*1.51*	0.93	1.00	*0.09*	-0.85	*2.79*	0.75	0.82	*0.06*	-1.07	*2.24*	0.83
2011	New Haven	KY	0.81	*0.08*	5.67	*3.02*	0.68	0.72	*0.08*	7.13	*2.89*	0.40	0.60	*0.09*	9.91	*3.51*	0.42
2009	Hopkinsville	KY	†	†	†	†	†	†	†	†	†	†	0.61	*0.10*	12.56	*4.00*	0.33
2011	Hopkinsville	KY	0.64	*0.06*	7.39	*3.10*	0.74	0.89	*0.05*	3.35	*1.84*	0.89	0.63	*0.05*	7.75	*2.51*	0.78
2009	Keiser	AR	0.58	*0.02*	2.09	*0.79*	0.92	†	†	†	†	†	†	†	†	†	†
2010	Keiser	AR	0.67	*0.03*	6.18	*1.36*	0.89	†	†	†	†	†	†	†	†	†	†
2011	Keiser	AR	0.44	*0.02*	0.24	*0.98*	0.90	†	†	†	†	†	†	†	†	†	†
2009	Weiner	AR	0.56	*0.03*	2.56	*1.08*	0.91	0.97	*0.01*	0.50	*0.38*	0.99	0.55	*0.02*	2.90	*1.08*	0.91
2010	Weiner	AR	0.60	*0.05*	14.73	*2.24*	0.71	†	†	†	†	†	†	†	†	†	†
2011	Weiner	AR	0.64	*0.03*	3.72	*1.17*	0.91	0.89	*0.03*	1.08	*0.81*	6.65	0.57	*0.04*	4.40	*1.53*	0.73
2010	Colt	AR	0.64	*0.04*	1.73	*2.07*	0.89	†	†	†	†	†	†	†	†	†	†
2011	Colt	AR	0.64	*0.03*	3.90	*1.21*	0.86	†	†	†	†	†	†	†	†	†	†
2009	Winnsboro	LA	1.59	*0.06*	-6.12	*1.90*	0.95	0.48	*0.05*	11.00	*2.76*	0.69	0.80	*0.08*	7.10	*2.88*	0.70
2010	Winnsboro	LA	0.91	*0.02*	0.78	*0.83*	0.99	0.85	*0.02*	1.66	*0.59*	0.98	0.78	*0.01*	1.98	*0.60*	0.99
2011	Winnsboro	LA	0.85	*0.01*	1.27	*0.51*	0.99	0.77	*0.02*	1.66	*0.62*	0.98	0.66	*0.02*	2.60	*0.65*	0.98
2009	St. Joseph	LA	0.73	*0.05*	1.55	*2.14*	0.85	0.81	*0.04*	0.82	*1.20*	0.93	0.57	*0.06*	2.80	*2.28*	0.73
2010	St. Joseph	LA	0.63	*0.05*	4.69	*2.09*	0.79	0.81	*0.04*	-1.77	*1.32*	0.91	0.52	*0.05*	1.62	*2.08*	0.76
2011	St. Joseph	LA	0.65	*0.03*	0.25	*1.94*	0.92	0.98	*0.02*	-0.09	*0.80*	0.97	0.64	*0.03*	-0.02	*1.66*	0.90
2010	Baton Rouge	LA	0.89	*0.03*	3.56	*1.56*	0.96	0.83	*0.03*	2.27	*1.46*	0.93	0.76	*0.03*	4.63	*1.35*	0.91
2011	Baton Rouge	LA	1.13	*0.03*	0.29	*1.36*	0.97	0.78	*0.03*	3.08	*1.43*	0.93	0.89	*0.04*	2.94	*1.68*	0.94
2010	Crowley	LA	0.92	*0.04*	0.32	*1.92*	0.94	0.92	*0.02*	2.12	*0.71*	0.99	0.86	*0.03*	2.03	*1.48*	0.94
2011	Crowley	LA	0.69	*0.02*	-3.70	*0.93*	0.95	0.96	*0.01*	-0.15	*0.12*	1.00	0.66	*0.02*	-3.71	*0.89*	0.95
Average Parameter Estimates			**0.81**	***0.01***	**2.51**	***0.38***	**0.92**	**0.89**	***0.02***	**2.20**	***0.35***	**0.92**	**0.74**	***0.01***	**3.80**	***0.46***	**0.89**

†Plant density data not captured at environment

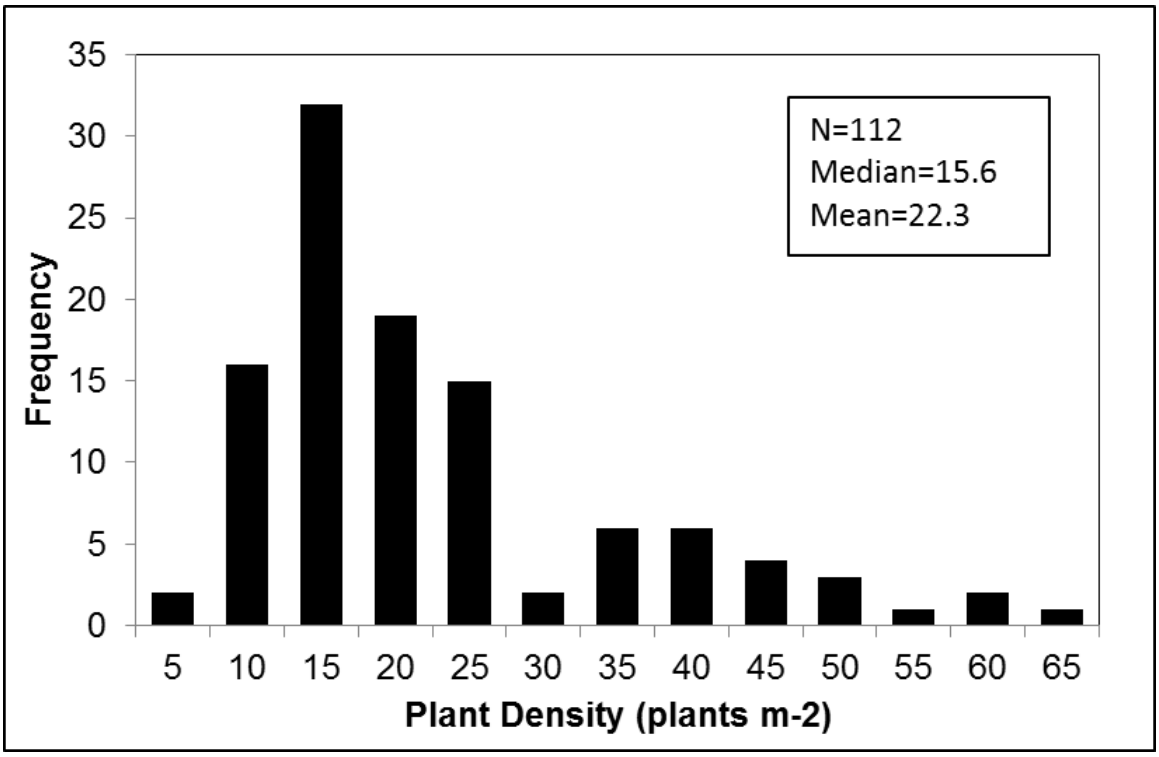


Figure 7. Distribution of plant density required to achieve 95% asymptotic yield across environments and cultivars. 112 locations are included in the distribution rather than the complete set of 118; 6 location*cultivar combinations did not converge for the β responsiveness parameter. Data points for three location by cultivar combinations at Keiser, AR which had plant densities greater than 95 plants m^{-2} are not displayed in the figure, but are represented in the median and mean calculations.

Literature cited

Ablett, G.R., W.D. Beversdorf, and V.A. Dirks. 1991. Row width and seeding rate performance of indeterminate, semideterminate, and determinate soybean. J. Prod. Agric. 4:391-395.

Anderson, L.R., and B.L. Vasilas. 1985. Effects of planting date on two soybean cultivars: Seasonal dry matter accumulation and seed yield. Crop Sci. 25:999-1004.

Bandillo, N., D. Jarquin, Q. Song, R. Nelson, P. Cregan, J. Specht, and A. Lorenz. 2015. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Gen. doi:10.3835/plantgenome2015.04.0024

Beuerlein, J.E. 1988. Yield of indeterminate and determinate semidwarf soybean for several planting dates, row spacings, and seeding rates. J. Prod. Agric. 1:300-303.

Bernard, R.L., C.R. Cremeens, R.L. Cooper, F.I. Collins, O.A. Krober, K.L. Athow, F.A. Laviolette, C.J. Coble, R.L. Nelson. 1998. Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000--IV (FC 01.547--PI 266.807). TBull USDA 1844. Comment: Data is in GRIN with evaluation name SOYBEAN.EVALUATION.1MN63

Board, J.E., and H. Modali. 2005. Dry matter accumulation predictors for optimal yield in soybean. Crop Sci. 45:1790-1799.

Bolon, Y.-T., B. Joseph, S.B. Cannon, M.A. Graham, B.W. Diers, A.D. Farmer, G.D. May, G.J. Muehlbauer, J.E. Specht, Z. J. Tu, N. Weeks, W.W. Xu, R.C. Shoemaker, and C.P. Vance. 2010. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Biology. 10:41.

Bolon, Y.-T., A.O. Stec, J.-M. Michno, J. Roessler, P.B. Bhaskar, L. Ries, A.A. Dobbels, B. W. Campbell, N.P. Young, J.E. Anderson, D.M. Grant, J.H. Orf, S.L. Naeve, G.J. Muehlbauer, C.P. Vance, and R.M. Stupar. 2014. Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics 198: 967-981.

Breene, W.M., S. Lin, L. Harman, and J.H. Orf. 1988. Protein and oil content of soybeans from different geographic locations. J. Am. Oil Chem. Soc. 65:1927-1931.

Brim, C.A., 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220

Brim, C.A. and J.W. Burton. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Broman, K.W., H. Wu, S. Sen, G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890

Brumm, T.J. and C.R. Hurburgh. 1990. Estimating the processed value of soybeans. J. Am. Oil Chem. Soc. 67: 302-307.

Brumm, T.J. and C.R. Hurburgh. 2006. Changes in long-term soybean compositional patterns. J. Am. Oil Chem. Soc. 83:981-983.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370-378.

Carpenter, A.C., and J.E. Board. 1997a. Branch yield components controlling yield stability across plant populations. Crop Sci. 37:885-891.

Carpenter, A.C., and J.E. Board. 1997b. Growth dynamics controlling soybean yield stability across plant populations. Crop Sci. 37:1520-1526.

Chen, G., and P. Wiatrak. 2011. Seeding rate effects on soybean height, yield, and economic return. Agron. J. 103:1301-1307.

Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053-1067.

Cober, E.R. and H.D. Voldeng. 2000. Developing high-protein, high-yield soybean populations and lines. Crop Sci. 40:39-42.

Conley, S.P., L. Abendroth, R. Elmore, E.P. Christmas, and M. Zarnstorff. 2008. Soybean seed yield and composition response to stand reduction at vegetative and reproductive stages. Agron. J. 100:1666-1669.

Concibido, V.C., B.W. Diers, and P.R. Arelli. 2004. A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci. 44:1121-1131.

Costa, J.A., E.S. Oplinger, and J.W. Pendleton. 1980. Response of soybean cultivars to planting patterns. Agron. J. 72:153-156.

Coulter, J. 2011. Yield and harvest considerations for frost damaged corn. Available at http://blog.lib.umn.edu/efans/cropnews/2011/09/yield-and-harvest-consideratio.html [verified 5 May 2014]. Minnesota Crop News.

Cox, W.J., and J.H. Cherney. 2011. Growth and yield responses of soybean to row spacing and seeding rate. Agron. J. 103:123-128.

DeBruin, J.L., and P. Pedersen. 2008a. Soybean yield response to planting date and seeding rate in the Upper Midwest. Agron. J. 100:696-703.

DeBruin, J.L. and P. Pedersen. 2008b. Effect of row spacing and seeding rate on soybean yield. Agron. J. 100:704-710.

Devlin, D.L., D.L. Fjell, J.P. Shroyer, W.B. Gordon, B.H. Marsh, L.D. Maddux, V.L. Martin, and S.R. Duncan. 1995. Row Spacing and seeding rates for soybean in low and high yielding environments. J. Prod. Agric. 8:215-222.

Diers, B.W., Keim, P., W.R. Fehr, R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theoret. Appl. Genet. 83: 608-612.

Dornbos, D.L. and R.E. Mullen. 1992. Soybean seed protein and oil contents and fatty-acid composition adjustments by drought and temperature. J. Am. Oil Chem. Soc. 69:228-231.

Dornbos, D.L., R.E. Mullen, and R.M. Shibles. 1989. Drought stress effects during seed fill on soybean seed-germination and vigor. Crop Sci. 29:476-480.

Eberhart, S.A. and W.A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6:36-40.

Duncan, W.G. 1986. Planting patterns and soybean yield. Crop Sci. 26:584-588.

Edwards, J.T. and L.C. Purcell. 2005. Soybean yield and biomass responses to increasing plant population among diverse maturity groups: I. Agronomic characteristics. Crop Sci 45:1770-1777.

Egli, D.B. 1988. Plant density and soybean yield. Crop Sci. 28:977-981.

Egli, D.B. and P.L. Cornelius. 2009. A regional analysis of the response of soybean yield to planting date. Agron. J. 101:330-335

Elmore, R.W. 1990. Soybean cultivar response to tillage systems and planting date. Agron. J. 82:69-73.

Elmore, R.W. 1991. Soybean cultivar response to planting rate and tillage. Agron. J. 83:829-832.

Elmore, R.W. 1998. Soybean cultivar responses to row spacing and seeding rates in rainfed and irrigated environments. J. Prod. Agric. 11:326-331.

Epler, M., and S. Staggenborg. 2008. Soybean yield and yield component response to plant density in narrow row systems. Online. Crop Management doi:10.1094/CM-2008-0925-01-RS.

Ethridge, W.J., Jr., D.A. Ashley, and J.M. Woodruff. 1989. Row spacing and plant population effects on yield components of soybean. Agron. J. 81:947-951.

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development, Special Report 80, Ames, IA, Ext. Service, Agric. Exp. Stn., Iowa State University.

Fehr, W.R., J.S. Burris, and D.F. Gilman. 1973. Soybean emergence under field conditions. Agron. J. 65:740-742

Fehr. W.R. Breeding for modified fatty acid composition in soybean. Crop Sci. 47:S72-S87.

Finlay, K.W. and G.N Wilkinson. 1963. The analysis of adaptation in a plant breeding program. Aust. J. Agric. Res. 14:742-752

Fox, C.M., T.R. Cary, A.L. Colgrove, E.D. Nafziger, J.S. Haudenshield, G.L. Hartman, J.E. Specht, and B.W. Diers. 2013. Crop Sci. 53:2473-2482.

Haley, C.S. and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315-324.

Ham, G.E., I.E. Liener, S.D. Evans, R.D. Frazier, and W.W. Nelson. 1975. Yield and composition of soybean seed as affected by N and S fertilization. Agron. J. 67:293-297.

Hamman, B., D.B. Egli, and G. Koning. 2002. Seed vigor, soilborne pathogens, preemergent growth, and soybean seedling emergence. Crop Sci. 42:451-457.

Hanson, W.D., R.C. Leffel, and R.W. Howell. 1961. Genetic analysis of energy production in the soybean. Crop Sci. 1:121-126.

Hartwig, E.E. and K. Hinson. 1972. Association between chemical composition of seed and seed yield of soybeans. Crop Sci. 12:829-830.

Heatherly, L.G. 1999. Early Soybean Production System (ESPS). In: L. G. Heatherly and H. F. Hodges, Eds., Soybean Production System in the Midsouth, CRC Press, Boca Raton, 1999, pp. 103-115.

Heatherly, L.G. and R.W. Elmore. 2004. Managing inputs for peak production. In: Boerma, R.H, and J.E. Specht (eds) Soybeans: Improvement, production, and uses, 3rd edn. American Society of Agronomy, Madison, WI, pp 451-536.

Hegstad, J.M., G. Bollero, and C.D. Nickell. 1999. Potential of using plant row yield trials to predict soybean yield. Crop Sci. 39:1671-1675.

Helms, T.C., C.R. Hurburgh, R.L. Lussenden, and D.A. Wited. 1990. Economic-analysis of increased protein and decreased yield due to delayed planting of soybean. J. Prod. Agric. 3:367-371.

Helms, T.C., E. Deckard, R.J. Goos, and J.W. Enz. 1996. Soybean seedling emergence influenced by days of soil water stress and soil temperature. Agron.J. 88:657-661.

Helms, T.C. and J.H. Orf. 1997. Protein , oil, and yield of soybean lines selected for increased protein. Crop Sci. 38:707-711.

Hurburgh, C.R. 1994. Long-term soybean composition patterns and their effect on processing. J. Am. Oil Chem. Soc. 71:1425-1427.

Hwang, E.-Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 15:1.

Hyatt, J., O. Wendroth, D.E. Egli, and D.M. TeKrony. 2007. Soil compaction and soybean seedling emergence. Crop Sci. 47:2495-2503.

Hyten, D.L., I.-Y. Choi, Q. Song, J.E. Specht, T.E. Carter Jr., R.C. Shoemaker, E.-Y. Hwang, L.K. Matukumalli, and P.E. Cregan. 2010. A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci. 50:960-968.

Johnson, H.W., H.F. Robinson, and R.E. Comstock. 1955. Genotypic and phenotypic correlations in soybeans and their implications in selection. Agron. J. 47:477-483.

Kratochvil, R.J., J.T. Pearce and M.R. Harrison, Jr. 2004. Row-spacing and seeding rate effects on glyphosate-resistant soybean for Mid-Atlantic production systems. Agron. J. 96:1029-1038.

Kravchenko, A.N., and D.G. Bullock. 2002. Spatial variability of soybean quality data as a function of field topography: I: Spatial data analysis. Crop Sci. 42:804-815.

Lambert, J.W. and B.W. Kennedy. 1975. Registration of 'Evans' Soybean. Crop Sci 15:735.

Lee, C.D. 2006. Reducing row widths to increase yield: Why it does not always work. Online. Crop Management doi:10.1094/CM-2006-0227-04-RV.

Lee, C.D., D.B. Egli, and D.M. TeKrony. 2008. Soybean response to plant population at early and late planting dates in the Mid-South. Agron. J. 100:971-976.

Lehman, W.F. and J.W. Lambert. 1960. Effects of spacing of soybean plants between and within rows on yield and its components. Agron. J. 52:84-86.

Lueschen, W.E. and D.R. Hicks. 1977. Influence of plant population on field performances of three soybean cultivars. Agron. J. 69:390-393.

Mahmoud, A.A., S.S. Natarajan, J.O. Bennett, T.P. Mawhinney, W.J. Weibold, and H.B. Krishnan. 2006. Effect of six decades of selective breeding on soybean protein composition and quality: A biochemical and molecular analysis. J. Agric. Food Chem. 54:3916-3922

Medic, J., C. Atkinson, and C.R. Hurburgh. 2014. Current knowledge in soybean composition. J. Am. Oil Chem. Soc. 91:363-384.

Mohring, J. and H.-P. Piepho. 2009. Comparison of weighting in two-stage analysis of plant breeding trials. Crop Sci. 49:1977-1988.

Nakasathien, S.I., D.W. Israel, R.F. Wison, and P. Kwanyuen. 2000. Regulation of seed protein concentration by supra-optimal nitrogen supply. Crop Sci. 40:1277-1284.

Nichols, D.M., K.D. Glover, S.R. Carlson, J.E. Specht, and B.W. Diers. 2006. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Sci. 46:834-839.

Oplinger, E.S., and B.D. Philbrook. 1992. Soybean planting date, row width, and seeding rate response in three tillage systems. J. Prod. Agric. 5:94-99.

Orf, J.H. and T.C. Helms. 1994. Selection to maximize gross value per hectare within three soybean populations. Crop Sci. 34:1163-1167.

Palmer, R.G., T.W. Pfeiffer, G.R. Buss, and T.C. Kilen. 2004. Qualitative Genetics. In: Boerma, R.H, and J.E. Specht (eds) Soybeans: Improvement, production, and uses, 3rd edn. American Society of Agronomy, Madison, WI, pp 137-234.

Pedersen, P., and J.G. Lauer. 2002. Influence of rotation sequence and tillage system on the optimum Corn and soybean plant population. Agron. J. 94:968-974.

Pedersen, P., and J.G. Lauer. 2003. Soybean agronomic response to management systems in the upper Midwest. Agron. J. 95:1146-1151.

Pedersen, P., and J.G. 2004. Response of soybean yield components to management system and planting date. Agron. J. 96:1372-1381.

Perkins, J.M., and J.L. Jinks. 1968. Environmental and genotype-environment components of variability. III. Multiple lines and crosses. Heredity 23:339-356.

Piper, E.L., and K.J. Boote. 1999. Temperature and cultivar effects of soybean seed oil and protein concentration. J. Am. Oil Chem. Soc. 76:1233-1241.

Pipolo, A.E., T.R. Sinclair, and G.M.S. Camara. 2004. Effects of temperature on oil and protein concentration in soybean seeds cultured in vitro. Ann. Appl. Biol. 144:71-76.

Ray, J.D., F.B. Fritschi, and L.G. Heatherly. 2006. Large applications of fertilizer N at planting affects seed protein and oil concentration and yield in the Early Soybean Production System. Field Crops Res. 99:67-74.

Rincker, K., R. Nelson, J. Specht, D. Sleper, T. Cary, S.R. Cianzio, S. Casteel, S. Conley, P. Chen, V. Davis, C. Fox, G. Graef, C. Godsey, D. Holshouser, G.-L. Jiang, S.K. Kantartzi, W. Kenworthy, C. Lee, R. Mian, L. McHale, S. Naeve, J. Orf, V. Poysa, W. Schapaugh, G. Channon, R. Uniatowski, D. Wang, and B. Diers. 2014. Crop Sci. 54:1419-1432.

Rotundo, J.L., J.E. Miller-Garvin, and S.L. Naeve. 2016. Regional and temporal variation in soybean seed protein and oil across the United States. Crop Sci. 56:797-808.

Rotundo, J.L., and M.E. Westgate. 2010. Rate and duration of seed component accumulation in water-stressed soybean. Crop Sci. 50:676-684.

Rowntree, S.C., J.J. Suhre, N.H. Weidenbenner, E.W. Wilson, V.M. Davis, S.L. Naeve, S.N. Casteel, B.W. Diers, P.D. Esker, J.E. Specht, and S.P. Conley. 2013. Genetic gain x management interactions in soybean: I. Planting date. Crop Sci. 53:1128-1138.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 40:1438-1444.

Sebern, N.A. and J.W. Lambert. 1984. Effects of stratification for percent protein in two soybean populations. Crop Sci. 24:225-228.

Setiyono, T.D., K.G. Cassman, J.E. Specht, A. Dobermann, A. Weiss, H. Yang, S.P. Conley, A.P. Robinson, P. Pedersen, and J.L. DeBruin. 2010. Simulation of soybean growth and yield in near-optimal growth conditions. Field Crops Res. 119:161-174

Shannon, J.G., J.R. Wilcox, and A.H. Probst. 1972. Estimated gains from selection for protein and yield in the F4 generation of six soybean populations. Crop Sci. 12:824-826.

Simpson, A.M., Jr., and J.R. Wilcox. 1983. Genetic and phenotypic associations of agronomic characteristics in four high protein soybean populations. Crop Sci. 23:1077-1081.

Song, Q., D.L. Hyten, J. Gaofeng, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2015. Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes| Genomes| Genetics 50(10):1999-2006.

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high density genotyping array for soybean. PLoS One. 8(1)e:54985.

Specht, J.E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, J.H. Orf, and K.G. Lark. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. 41:493-509.

Suhre, J.J., N.H. Weidenbenner, S.C. Rowntree, E.W. Wilson, S.L. Naeve, S.P. Conley, S.N. Casteel, B.W. Diers, P.D. Esker, J.E. Specht, and V.M. Davis. 2014. Soybean yield portioning changes revealed by genetic gain and seeding rate interactions. Agron. J.106:1631-1642.

Thorne, J.C., and W.R. Fehr. 1970. Incorporation of high-protein, exotic germplasm into soybean populations by 2- and 3-way crosses. Crop Sci. 10:652-655.

Torrion, J., T.D. Setiyono, K. Cassman, and J. Specht. 2011. Soybean phenology simulation in the North-Central United States. Agron. J. 106:1661-1667.

Unander, D.W., J.H. Orf, and J.W. Lambert. 1986. Early season cold tolerance in soybean. Crop Sci. 26:676-680.

Updaw, N.J. and T.E. Nichols, Jr. 1980. Pricing of soybean on the basis of chemical constituents. P. 781-799. In F.T. Corbin (ed.) Proc.World Soybean Res. Conf., 2nd, Raleigh, NC. 26-29 Mar. 1979. Westview Press, Boulder, CO.

USDA-ARS National Genetic Resources Program. 2016. Germplasm Resources Information Network (GRIN) [Online Database]. Glycine max (L.) Merr. PI 153296. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1136705. National Germplasm Resources Laboratory, Beltsville, MD. (20 April 2016)

Vaughn, J.N., R.L. Nelson, Q. Song, P.B. Cregan, and Z. Li. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3 4:2283-2294.

Vetch, J.A., G.W. Randall, and J.A. Lamb. 2007. Corn and soybean production as affected by tillage systems. Agron. J. 99:952-959.

Voldeng, H.D., E.R. Cober, D.J. Hume, C. Gillard, and M.J. Morrison. 1997. Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. Crop Sci. 37:428-431.

Walker, E.R., A. Mengistu, N. Bellaloui, C.H. Koger, R.K. Roberts, and J.A. Larson. 2010. Plant population and row-spacing effects on maturity group III soybean. Agron. J. 102:821-826

Weber, C.R., R.M. Shibles, and D.E. Byth. 1966. Effect of plant population and row spacing on soybean development and production. Agron. J. 58: 99-102.

Wehrmann, V.K., W.R. Fehr, S.R. Cianzio, and J.F. Cavins. 1987. Transfer of high seed protein to high-yielding soybean cultivars. Crop Sci. 27:927-931.

Whitaker, D., E.R. Williams and J.A. John. 2006. *CycDesigN: A Package for the Computer Generation of Experimental Designs*. CycSoftware Ltd., Hamilton.

Wiggans, R.G. 1939. The influence of space and arrangement on the production of soybean plants. J. Am. Soc. Agron. 31:314-321.

Williams, E.R., J.A. John and D. Whitaker. 2006. Construction of resolvable spatial row-column designs. Biometrics 62: 103-8

Wison, R.F. 2004. Seed composition. In: Boerma, R.H, and J.E. Specht (eds) Soybeans: Improvement, production, and uses, 3rd edn. American Society of Agronomy, Madison, WI, pp 621-669.

Wison, E.W., S.C. Rowntree, J.J. Suhre, N.H. Weidenbenner, S.P. Conley, V.M. Davis, B.W. Diers, P.D. Esker, S.L. Naeve, J.E. Specht, and S.N. Casteel. 2013. Genetic

gain x management interactions in soybean: II. Nitrogen utilization. Crop Sci. 54:340-348.

Wilcox, J.R. and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1536-1540.

Wilcox. J.R., and E.M. Fankenberger. 1987. Indeterminate and determinate soybean responses to planting date. Agron. J. 79:1074-1078.

Wilcox, J.R. and Z. Guodong. 1997. Relationships between seed yield and seed protein in determinate and indeterminate soybean populations. Crop Sci. 37:361-364.

Wilcox, J.R. and R.M Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Yaklich, R.W., M.M. Kulik, and C.S. Garrison. 1979. Evaluation of vigor in soybean seeds: Influences of date of planting and soil type on emergence, stand, and yield. Crop Sci. 19:242-246.

Yaklich, R.W., B. Vinyard, M. Camp, and S. Douglass. 2002. Analysis of protein and oil from soybean northern and southern region uniform tests. Crop Sci. 42:1504-1515.